Automated analysis of sewer CCTV surveys

Submitted by Joshua Myrans, to the University of Exeter as a thesis for the degree of Doctor of Philosophy in Water Informatics Engineering, October 2018.

I certify that all material in this thesis which is not my own work has been identified and that any material that has previously been submitted and approved for the award of a degree by this or any other University has been acknowledged.

……

Abstract

Sewers across the globe must be regularly inspected to ensure their smooth running and effective maintenance. Furthermore, surveys are often performed reactively, often diagnosing suspected faults within a network. Almost all surveys in the UK and abroad are performed using closed circuit television (CCTV) cameras where pipes are too small for manual inspection. As such vast quantities of footage are recorded by surveying teams on a daily basis. This footage is currently analysed manually, requiring a trained engineer to watch through its entirety, annotating potential faults. This thesis examines methods of improving this labelling process, implementing various machine learning and image processing techniques to automate this procedure. The thesis presents two distinct methodologies: the first for the detection of faults, using only raw CCTV footage, whilst the second identifies the type of a detected fault according to the *Manual of Sewer Condition Classification.*

The fault detection methodology identifies the presence of a fault within a CCTV image. The methodology calculates a GIST feature descriptor for each video frame, before utilising a Random Forest classifier, to predict the presence of a fault. The basic methodology was further refined with the inclusion of smoothing, to eliminate isolated inconstancies, and stacking to intuitively combine the results of multiple machine learning classifiers. The final methodology achieved a detection accuracy of 86% on unseen real-life data from the UK.

The fault classification methodology identifies the fault type in images, where faults have been previously detected using the above technique. The tool again calculates a frame's GIST descriptor before applying multiple Random Forest classifiers in a '1 vs all' architecture to predict the type of a given fault. This architecture allowed for comparative classifications and later enabled the identification of multiple faults within a single frame. The methodology achieved a peak accuracy of 74% when classifying faults well represented by the dataset (at least 100 examples). Furthermore, when including multi-label functionality, the tool achieved an accuracy of 67% across all fault types.

Both methods have been developed to be holistic and practical, utilising only industry standard CCTV footage and generalising well across all types of sewer system (size, shape and material). Furthermore, as both methodologies rely on the same feature descriptor, they integrate well to form a methodology that could be applied in real time. As such the thesis also explores the practical implications of creating a detection support tool capable of integration with current working practices.

In combination both methodologies and their additions present a unique contribution to the field of automated sewer surveying. Achieving competitive accuracies with a streamlined methodology, the technology shows promise for future application in industry, greatly increasing the speed, accuracy and consistency of CCTV sewer surveys.

Acknowledgements

I would like to thank these people, without whom this work would not have been possible:

Prof. Zoran Kapelan

for being my supervisor over the last four years, guiding me through the world of academic engineering, for always encouraging me to push myself and share my work with the world, and most importantly for doing all the above with great patience.

Prof. Richard Everson

for inspiring my interest in machine learning many years ago and still choosing to supervise me through this course, for the patience and understanding shown throughout the many technical aspects of my work, and for doing so with great spirit and humour.

Julian Britton

for generously providing the data and industrial insight necessary for this work to succeed.

Alicia Joyce

for her constant love support and encouragement.

My parents **Julian** and **Julia Myrans**

for their unconditional love, guidance and support, making me the person I am today.

Contents

List of Tables

List of Figures

Chapter 1: Introduction

Sewer surveys are often expensive and laborious tasks, requiring the analysis of large volumes of recorded CCTV footage. Alongside the rapid development of new machine learning and image processing technologies, the possibility to automate the analysis process is becoming increasingly viable. This chapter introduces the motivations behind the thesis, discussing current surveying practices and its issues. A solution to these issues is presented, discussing the improvements an automated CCTV analysis tool could provide. This leads to an outline of the aims of this work and the novel contributions it presents. Finally, an overview of the thesis' subsequent chapters is laid out alongside work published in relation to this thesis.

1.1 Current Sewer Surveying Practice

Water companies across the globe own thousands of kilometres of underground sewer pipes. Although located underground and hence not easily accessible, these pipes are pivotal to the smooth operation of daily life, as such the networks undergo regular inspections. These inspections, performed using Closed Circuit Television (CCTV) cameras, assess the condition of a pipe. A typical survey will identify the location, type and severity of potential faults, compiling the information into a report for ease of access. This information is vital for the effective maintenance of the entire network and enables a company to prioritise pipe refurbishment.

As most pipes are too small for human access, CCTV cameras remotely traverse the network, using a variety of techniques. The two most commonly applied technologies include a PIG (Pipe Inspection Gadget) and a system of push rods. A PIG is a small device, mounted with a camera and light, that is remotely guided through the network. Much like a remote-control car, the PIG drives through a pipe recording it's interior. Alternatively, the simpler push-rod system can achieve similar results. This recording technique forces a camera through the pipe using semi-rigid cables, again recording the pipe's interior. In both cases the camera enters the network via a manhole and has a good range, relying solely on the length of available cable or number of available push rods. Given the correct equipment both techniques can provide a live feed from the camera to the

surveyor, enabling some level of control over the camera's pace and direction. This control is further improved by some PIGs, which have inbuilt pan, tilt and zoom capabilities, enabling a surveyor to thoroughly investigate potential faults. A PIG may also be equipped with additional technologies to supplement the camera feed. These technologies may include: laser profilers, which can accurately measure the internal dimensions of a pipe (IBAK, 2018) and 3D cameras, capable of recording 360-degree panoramic footage of a pipe (Duran, Althoefer, & Seneviratne, 2002). Unfortunately, these technologies are expensive and not currently used in common practice. As such these supplementary technologies are not relied upon in any of these studies but may be considered for future investigations.

CCTV footage collected for a pipe survey is annotated with key details, including details about any encountered fault. These annotations are used to compile an overall survey report, used by a water company to assess the condition of the pipe. In the UK annotations are made according to the *WRc Manual of Sewer Condition Classification* (WRc plc, 2013), and can be written 'offline' or 'online'. 'Offline' annotation is performed once the entire survey has been recorded by a surveyor in the field. The footage is often returned to an office, where a technician re-watches the footage, annotating faults with the relevant details. Conversely, 'online' annotation can be performed during the collection of the footage when a live feed is available. A surveyor watches the footage as it is collected, annotating any encountered faults. Online annotation lends itself to the use of a PIG, where the surveyor has full control of the camera and can stop the camera whilst making notes. Regardless of the chosen approach specialised software, such as *WinCan* (CD Lab AG, 2018) is often used to perform the annotation and compile the report.

1.2 The Problem

Although current techniques have proved effective and are widely applied across the globe, they are not without issues. Firstly, surveying can be extremely slow and tedious work. If CCTV footage is analysed 'offline' the entire duration of a survey (hours or days of footage) must be re-watched for annotation. Alternatively, when footage is analysed 'online' the survey is constantly paused, allowing the operator time to make the necessary annotations every time a fault

occurs. Stacking these inefficiencies on top of the difficulties surveyors face when locating pipes and finding suitable access points make surveying an extremely laborious task.

From a water company's perspective, the slow nature of surveys makes them expensive. Equipment and staff make up a sizeable proportion of a survey's expense, but often a company is also charged for accessing their assets. A local council may fine the company for causing disruption to local infrastructure caused when accessing sewers, or additional equipment and expertise may be required to survey particularly difficult to access assets. All of these above costs are often charged at an hourly rate, in turn costing the company more the longer a survey takes.

Finally, due to human nature and the monotony of sewer surveys it is common for faults to be missed or mislabelled. Like all humans, surveyors make mistakes, especially when annotating large volumes of homogeneous CCTV footage. In addition, faults are often labelled subjectively, with every surveyor having a slightly different interpretation of the surveying standards. Surveyors often label footage at differing distances from a fault, while other surveyors may only record faults if they are judged to be above a certain severity. Given human nature and the differences in training and company custom, these differences are unavoidable and have been well documented by van Der Steen et al. (2014) alongside other inconsistencies in current surveying practices.

1.3 An Automated Methodology and Tool

The work presented within this thesis aims to improve the accuracy, consistency and efficiency of current sewer surveying practice by automating much of the sewer survey annotation process. Automating aspects of the annotation process directly addresses many issues with current surveying practices.

An automated methodology would increase the speed at which surveys can be annotated, allowing surveys to be annotated in real-time, without the need to slow surveys for human analysis. Alternatively, automated annotation could be performed 'offline' outside of office hours, on unused computers, utilising idle time and resources. A faster, more efficient survey reduces its associated costs, enabling water companies to survey a larger proportion of their sewer network on

a regular basis. Indirectly this could save money, as more faults will be identified and addressed before they develop into much larger, more expensive problems.

On the other hand, a single standardised analysis tool would eliminate most human errors, never getting tired or bored e.g. Missing an important fault may result in a sewer blockage and a follow-on costly pollution or flooding incident. Similarly, a homogenous tool would eliminate the subjectivity, which is common in current sewer surveys. Correctly trained the tool could follow standard surveying guidelines, consistently labelling the fault location and providing uniform labels for each fault type. All this would increase the reliability of fault annotation.

In response to these challenges, the methodologies demonstrated in this thesis aim to integrate with current sewer surveying practice. All techniques are developed and tested using only industry standard CCTV footage, eschewing any retrofits or non-standard surveying hardware. In the same sense, the technology has been developed to be used as a detection support tool, working alongside a surveyor and reporting its findings for human confirmation. This redundancy would ensure high quality surveys, and leaves open the option for full automation, which could be applied as the methodologies gain the confidence and trust of surveyors.

1.4 Research Questions and Aims

1.4.1 Research Questions

The work in this thesis addresses the following key research questions:

1. Can faults be automatically detected in CCTV sewer surveys, using a combination of Image processing and machine learning techniques? If so can this be done in an efficient manner, which generalises across the many pipe and fault types present in sewer systems?

2. Given an image containing a fault, can its type be automatically identified? Could a developed method be easily integrated into a fault detection system?

3. How could developed technologies be effectively integrated with current industry practices? Could this integration be flexible, able to be implemented within both ‘online’ and ‘offline’ practices?

1.4.2 Aim and Objectives

The overall aim of the work in this thesis is to develop and demonstrate a collection of tools for the practical automation of CCTV sewer survey analysis. The specific objectives are as follows:

1. Development of a method capable of automated detection of faults in sewer pipes using only industry standard CCTV footage. This system should work in a holistic manner, capable of generalising to all environments (pipes and faults), currently surveyed using CCTV footage. The proposed detection methodology must be flexible to the needs of a field technician, and able to easily integrate with existing surveying practices.
2. Development of a method for the automated identification of fault types in sewers. Again, this should utilise only industry standard CCTV footage and generalise well to all surveyed environments. The technique needs to be flexible to the requirements of a field technician and ideally integrate with the automated fault detection system. This should result in a single streamlined process for the complete automated analysis of CCTV sewer surveys.
3. Test, validate and demonstrate the above new methodologies on real-life data. These demonstrations should reflect real life practices, evaluating the accuracy and efficiency of each technique.
4. Develop guidelines for the practical implementation of a detection support tool encapsulating the new methodologies. The guidelines should suggest best practice for both fault detection and classification, defining recommended configurations for automated analysis. This should be accompanied by example mock interfaces detailing options for future detection support tools.

1.5 Thesis Contributions

The following contributions are novel to this thesis:

- A new fault detection methodology, capable of identifying the presence of a fault in an image, as defined by the *Manual of Sewer Condition Classification* (WRc plc, 2013). This process applies GIST feature descriptors and Random Forests to identify faults in single frames of CCTV footage. These frames could be isolated images or extracted and processed in sequences from continuous CCTV footage.
- A collection of new tools and techniques for the optimisation and application of the fault detection methodology to continuous CCTV footage. The most notable additions include:
    - Smoothing, improving the accuracy of the fault detection method by incorporating information from neighbouring frames when detecting faults in continuous CCTV footage.
    - Stacking, intuitively combining the classifications of multiple machine learning classifiers to further improve the accuracy of the fault detection.
- A new fault type identification methodology, which identifies the categories of faults within an image, as defined by the *Manual of Sewer Condition Classification* (WRc plc, 2013). This methodology extends the already existing fault detection methodology, utilising an ensemble of Random Forest classifiers to identify the most probable fault classes within a frame.
- New guidelines to produce a suitable detection support tool, combining the above technologies, to assist technicians in the field. This tool includes advice for the set up and application of the developed tools as well as practical mock-ups of end-user software.

1.6 Thesis Overview

The remainder of this thesis is structured as follows:

Chapter 2: Background

Begins by introducing previous work in the field of automated CCTV fault analysis, working chronologically through methods and developments. Given its prevalence in fields outside of the water sector, autonomous crack detection is discussed in detail. This leads to an overall discussion of the object recognition techniques utilised by the latest work, and the feature descriptors which underpin many of the most successful approaches.

Chapter 3: Data

Introduces the datasets used throughout this thesis. The chapter begins by identifying the sources of data, and the collection methods implemented throughout the UK water industry. Next, the core Wessex Water dataset, containing over 50 hours of sewer surveys, is described. This discussion includes a description of the extraction and labelling process used to generate the core training and testing sets. Finally, this library of still images is examined, identifying the distributions of faults and pipe specifications prevalent throughout the dataset.

Chapter 4: Fault Detection Method

The first methodological chapter, this introduces the automated fault detection system. It begins by describing the method's four stages: 'Frame Extraction', 'Pre-processing', 'Feature Extraction' and 'Classification', providing detailed descriptions of the GIST feature descriptor and Random Forest (RF) classifier. The chapter continues to demonstrate the new methodology on still images and continuous CCTV footage, achieving a peak accuracy of 85%.

Chapter 5: Extended Fault Detection Method

This chapter continues to explore a collection of additional techniques and tools useful to the automated fault detection methodology. These include smoothing, stacking and a collection of alternative machine learning classifiers. Smoothing applies a Hidden Markov Model (HMM) and Order Oblivious Filtering (also known as windowing) to improve predictions in continuous footage by including

information from neighbouring frames. On the other hand, stacking attempts to combine the predictions of multiple machine learning classifiers using a secondary stacking classifier. This improves the accuracy and reliability of the fault detection method, by combining the strengths of multiple different classifiers. Finally, a collection of smaller techniques is discussed, including the location of a fault within an image and the application of anomaly detection tools.

Chapter 6: Fault Classification Method

Chapter 6 describes and demonstrates the new fault classification method. As before, the methodology is described, before application to both still images and continuous CCTV footage. With the overall aim of integrating both fault detection and classification processes, the two systems share ‘Frame Extraction’, ‘Pre-processing’ and ‘Feature Extraction’ stages. The key difference between detection and classification lies in the classification, where an ensemble of Random Forest (RF) classifiers are applied in a ‘1 vs all’ architecture to achieve a peak classification rate of 74% on video sequences.

Chapter 7: A Detection Support Tool

The thesis’ seventh chapter proposes the best practices for implementation of both fault detection and classification methodologies in practice. It identifies stages within current practice, where automated analysis software could easily integrate. The chapter also notes important considerations when implementing the methods in practice, including labelling practices and workflows for combining detection and classification. Finally, the chapter offers mock interfaces for a detection support tool. These cover both online and offline annotation, identifying key features that could improve the efficiency of surveying.

Chapter 8: Conclusion

The final chapter summarises all the material developed and discussed throughout the thesis. This includes a summary of the results and developments presented in each chapter, before identifying the key conclusions of the work. Finally, the potential for further research and investigation is discussed. These include the identification of areas within the current methods for further development and testing, alongside whole new methodologies for the further automation of the survey annotation process.

1.7 Published Work

Each of the methodological chapters of this thesis can be attributed to work published over the course of the research. These publications, along with their respective chapter are detailed below in Table 1.1.

Table 1.1 *Mapping between thesis questions, objectives, chapters and publications.*

Thesis chapter	Research question	Objective	Publication
4	1	1, 3	(Myrans, Kapelan, & Everson, 2016a) (Myrans J. , Kapelan, Everson, & Britton, 2016b) (Myrans, Everson, & Kapelan, 2018b)
5	1	1, 3	(Myrans, Kapelan, & Everson, 2017) (Myrans, Kapelan, & Everson, 2018a) (Myrans, Everson, & Kapelan, 2018b) (Myrans, Kapelan, & Everson, 2018e)
6	2	2, 3	(Myrans, Everson, & Kapelan, 2018c) (Myrans, Kapelan, & Everson, 2018d)
7	3	4	-

Journal Papers

Myrans, J., Kapelan, Z. and Everson, R., 2018a. Combining classifiers to detect faults in wastewater networks. Water Science and Technology, 77(9), pp.2184-2189.

Myrans, J., Everson, R. & Kapelan, Z., 2018b. Automated detection of faults in sewers using CCTV image sequences. *Automation in Constrauction,* Volume 95, pp. 64-71.

Myrans, J., Everson, R. & Kapelan, Z., 2018c. Automated detection of fault types in CCTV surveys. *Journal of Hydroinformatics. (Accepted for publication)*.

Conference Papers

Myrans, J., Kapelan, Z. & Everson, R., 2016a. Automated detection of faults in wastewtaer pipes from CCTV footage by using random forests. Procedia Engineering, Volume 154, pp. 36-41.

Myrans, J., Kapelan, Z., Everson, R. and Britton, J., 2016b. Using support vector machines to identify faults in sewer pipes from CCTV surveys. International Computing and Control in the Water Industry Conference (CCWI) 2016. Electronic proceedings.

Myrans, J., Kapelan, Z. & Everson, R., 2017. Automatic detection of sewer faults using continuous CCTV footage. Sheffield, UK, International Computing and Control in the Water Industry Conference (CCWI) 2017. Electronic proceedings.

Myrans, J., Kapelan, Z. & Everson, R., 2018d. Automatic identification of sewer fault types using CCTV footage. Palermo, Italy, 13th International Hydroinformatics conference (HIC) 2018. Electronic proceedings.

Myrans, J., Kapelan, Z. & Everson, R., 2018e. Using automatic anomaly detection to identify faults in sewers. Kingston, Canada, International Computing and Control in the Water Industry Conference (CCWI) 2018. WDSA/CCWI Joint Conference Proceedings.

Chapter 2: Background

Automated sewer survey analysis has a rich and innovative background, combining machine learning and image processing technologies for the past 20 years. Over this period, CCTV recordings have remained the foundation of surveying practices, allowing surveyors to examine the otherwise inaccessible interior of many pipes. However, the practice has constantly evolved, applying the latest software and recording technologies to regularly improve the quality of surveys. In combination with advances in machine learning techniques, many new methodologies for the automatic analysis of sewer surveys become increasingly viable. This chapter provides an overview of automated fault detection tools that have been previously developed and the technologies key to the success of recent works. From this literature review the chapter continues to clearly define the novelty of the work presented in the remainder of this thesis.

2.1 Early Work

The earliest work on automated sewer survey analysis was performed by Moselhi & Shebab-Eldeen (1999), who attempted to create a generic fault detection methodology and software capable of detecting all types of faults. This generic methodology applied automated detection before video footage was commonly recorded in a digital format. This meant frames had to be 'grabbed' from the analogue video tapes before processing. In order to detect faults, Moselhi & Shebab-Eldeen identified edges in smoothed frames, using a gradient based edge detector similar to the Sobel edge detector (Gao, Zhang, Yang, & Liu, 2010). These edges were used to segment the image, before applying a single layer Artificial Neural Network (ANN) (Yegnanarayana, 2009) to classify each segment. This technique was demonstrated on only a single CCTV image, containing a collection of cracks. The technique successfully located 8 cracks, identifying the size and shape of each. Moselhi & Shebab-Eldeen (2000) continued this work, utilising the same edge detection techniques in combination with a 3-layer ANN. With a single hidden layer, the ANN was trained using back propagation (Yegnanarayana, 2009) to identify four fault types (cracks, displaced joints, spalling and a reduction in cross sectional area). Training the ANN on 878 individual feature vectors, the methodology achieved a 98% accuracy, correctly

classifying 214 of the 218 unseen test features. Similarly, Chae & Abraham (2001) applied ANN to the recordings of a new sewer scanning device. The technique, known as 'Sewer Scanning and Evaluation Technology' (SSET), used a rotating camera to map the surface of the pipe's interior. Almost 20 years later, the technology isn't commonplace in industry, although it does allow for effective inspection of a pipe's surface. Chae & Abraham (2001) enhanced the recorded surface maps using a gradient based edge detector, before identifying faults using fuzzy logic and a collection of small ANNs. The methodology was demonstrated on 172 unseen images, correctly identifying cracks and displaced joints in 83% of cases.

Although dated, these methodologies began to explicitly define the goal of autonomous fault detection for sewer surveys. Even though surveys were not recorded digitally, modern computing techniques were applied to converted images. Many of the techniques applied by these studies have since been applied widely within the field, most notably: edge detection, fuzzy logic and artificial neural networks. As such, the influence of this work cannot go unrecognised.

2.2 Crack Detection

Cracks (and fractures) are some of the most common sewer faults, whether they be longitudinal, spiral or circumferential. As such it is unsurprising that automated crack detection has a rich history, especially as cracks are a problem in many other engineering sectors. Oullette et al. (2004) demonstrated one of the earliest crack detection techniques, applying a multi-layered ANN to CCTV survey images. The neural network's structure was adapted to work from the raw images, processing 2D arrays of pixel values, whilst the ANNs internal weights were learnt using a Genetic Algorithm (GA). A GA is a population-based optimisation technique, based loosely around the theory of evolution (Deb, Pratap, Agarwal, & Meyarivan, 2002). This methodology achieved an accuracy of 92% when applied to a library of 100 images, where each image was cut from a frame to contain only a crack. This application of an ANN to raw images demonstrated the flexibility of machine learning technologies, capable of performing well on complex datasets. However, in hindsight it is clear that the application of a feature descriptor may have made the ANN's calibration process much simpler. Again, using an ANN, Sinha & Feiguth (2006) applied fuzzy logic

to identify cracks. Potential faults are identified by comparing light intensity values to a predetermined threshold. Using a process of image segmentation, each potential fault is cropped from the original frame, before being classified by an ANN. The output of the ANN is interpreted using fuzzy logic (Klir & Yuan, 1995), incorporating expert knowledge to improve the final prediction. Given 500 example images (60% training, 40% testing), containing a mix of cracks, laterals, holes and displaced joints, the methodology correctly identified 95% of all cracks, often including their subtype. Sinha & Feiguth were among the first to combine fuzzy logic with ANNs for this problem, incorporating expert knowledge to improve the decision-making process. Applied to segmented images, this methodology would have the added benefits of locating a fault within an image, if costing more processing time.

Chaki & Chattopadhyay (2010) eschewed the machine learning techniques applied by other classifiers, to produce a semi-analytical approach to the problem. Objects within an image were identified by examining variations in light intensities, before a defined shape and compactness factor were applied to predict the presence of a crack. Thresholds for these factors were defined using fuzzy logic. In small scale tests, the technique was applied to five cracks over 100 contiguous frames, correctly identifying the cracks in 97-100% of frames. Although the lack of a complex machine learning methodology is refreshing, this analytical approach was only demonstrated on five faults. Due to this small-scale demonstration, it is hard to tell if the method would generalise well across the large variety of cracks found in sewers.

Khalifa et al. (2014) produced a very similar analytical approach to Chaki & Chattopadhyay (2010). Khalifa et al. applied a Canny edge detector to define objects within a frame, before using its length and width to identify the presence of a crack. However, unlike Chaki & Chattopadhyay, this technique was demonstrated on 101 examples of cracks, with an accuracy of 90%. This affirms the techniques ability to generalise, however its ability to distinguish cracks from other fault types or normal pipe is not discussed. Halfawy & Hengmeechai (2013) also developed an analytical approach to the crack detection problem, utilising multiple image processing techniques. As per the previous techniques, Halfawy & Hengmeechai applied a Canny edge detector to identify objects of interest, before applying directional and circular filters to identify the presence of a crack.

Unlike previous analytical techniques, Halfawy & Hengmeechai considered the ovality of a crack, utilising the Hough transform to detect circular patterns, enabling the detection of longitudinal, circumferential and combined cracks. Again, the methodology was applied to a set of 100 frames, although unlike previous experiments 50% of the images did not contain a crack. This more realistic experiment demonstrated a promising accuracy of 85% and appears to be the most applicable and best demonstrated analytical technique.

Crack detection is not unique to sewer surveying and can be applied to all aspects of civil engineering, including other pipe, tunnels and bridges. Jahanshahi & Marsi (2012) applied 3D scene reconstruction to identify cracks in civil infrastructure. By matching Scale-Invariant Feature Transform (SIFT) points between 2D frames, Jahanshahi & Marsi reconstruct a detailed 3D scene. These scenes are processed by a machine learning classifier to detect the presence and size of cracks. Applied to a synthesized dataset of 420 images (containing 220 faults), the methodology achieved an accuracy of 78% when using a Support Vector Machine (SVM) and 80% when using a Neural Network.

Conversely Chen et al. (2017) developed a methodology to detect cracks in metallic cooling pipes for power plants. The methodology extracted Local Binary Pattern (LBP) features from each frame, before identifying the presence of a crack using a SVM. Although a more controlled environment than sewers, the technique was demonstrated on ~6000 images (~2000 containing faults), achieving a 76% accuracy. In addition, the methodology took around 1.9 seconds to analyse a single frame, making it one of the faster documented techniques. Both methodologies are impressive, using modern technologies and computing paradigms to achieve convincing results. Furthermore, the methods are demonstrated on substantial datasets, containing a mix of normal and faulty images. However, these techniques were developed for environments that are much more strictly controlled than a combined sewer system. For this reason, the techniques may struggle when applied to sewer surveys.

2.3 Sewer Fault Detection

Unfortunately for surveyors, sewer pipes contain a multitude of fault types varying greatly in shape, size and location. As such, specialised crack detectors are not as practical as they may seem, and more generic fault detection techniques, covering multiple fault types have been developed. One of the earliest generic fault detection techniques was developed by Duran et al. (2003), who suggested retrofitting CCTV cameras with a laser profiler. Although expensive, this device can accurately map the internal pipe walls, greatly improving the survey's data quality. Duran et al. used these readings as inputs for an ANN, to identify multiple surface defects. In lab-based experiments, where holes were made in clean pipes, the technique achieved 80-100% accuracy, demonstrating potential for further development. As such Duran et al. (2007) continued to develop the methodology, developing three ANNs which were combined to predict the presence of a fault. This proved even more effective detecting holes, cracks, displaced joints and obstacles with 80–100% accuracy. Although both studies show promise for the technology, laser profilers are still not widely used for sewer surveys (10 years later), in combination with the lack of 'real-world' testing, this makes the methodology impractical for application in industry.

Ahray et al. (2007a) developed an anomaly detection technique using a collection of image processing tools. The developed technique first extracted a panoramic region of interest containing only the immediate pipe wall. This image was then filtered, to reduce noise and its edges were detected using the Sobel method. The processed panoramic image is finally compared to an exemplar image of a pipe wall containing no faults. If the two images are sufficiently different a fault is flagged. When demonstrated on a dataset of 253 images (147 containing faults of multiple types) the methodology correctly identified the presence of a fault 84% of the time. As an initial anomaly detection technology, this appears to perform well, demonstrated on a reasonably sized dataset. However, it is often impractical to only consider the pipe wall whilst surveying, faults can appear in many different locations. In addition, the technology would likely require a larger set of exemplar images than reported to fully represent the diversity in shape, size and material of modern sewer networks.

Continuing the theme of anomaly detection, Guo et al. (2009a) utilised frame differencing to detect sudden changes in a pipe. If an individual frame is significantly different to its predecessor, the methodology assumes the presence of a fault, being flagged for additional investigation. Although this technique is unable to provide specific details, it was demonstrated with an accuracy of 84% on a dataset of 103 CCTV frames (57 contained faults). Guo et al. (2009b) continued to refine this technique, calculating and comparing frame's SIFT (Scale-Invariant Feature Transform) features when determining the difference between two frames. SIFT features enable the comparison of key, dynamically determined features, focussing each comparison. This improvement was clearly demonstrated on a set of 192 frames (containing 44 faults) with a 92% accuracy. Overall, these methodologies would perform extremely well as a screening tool for CCTV survey, potentially able to cut large chunks of normal pipe from the analysis process. The only notable drawback could be the use of a SIFT feature descriptor, as in their non-distributed form they can be computationally expensive to calculate (see Section 2.5.2 for additional details on SIFT).

Thinking outside the box, Halfawy & Hengmeechai (2014a) identified faults using the motion of a camera. Assuming the use of a remotely operated PIG, Halfawy & Hengmeechai calculated the optical flow of every frame, identifying regions of interest when the camera pans or tilts. By tracking the operator's motions, the technique can further investigate specific regions of interest, segmenting faults within a frame. When applied to a collection of 350 images, the methodology correctly identified regions of interest 96% of the time. Although interesting, this work relies heavily on the use of pan/tilt/zoom features on a PIG and the operators' motions. This makes it useful as a screening tool for offline analysis but would be of no use in push-rod or fully automated systems. Halfawy & Hengmeechai (2014b) continued their work, developing an alternative fault detection methodology. This methodology calculates every frame's Histogram of Oriented Gradients (HOG) features, which are processed by an SVM classifier to detect faults. Applied to examples of root intrusions in a set of 1000 images (~50% containing faults) the technique proved effective achieving an accuracy of 86%. Unlike the previous methodology, this appears to be practical and applicable to most surveying setups, the only major drawback of the work being

the lack of diversity of fault types used to demonstrate the method (using only root intrusions).

Finally, and most recently, Hawari et al. (2018) demonstrated a methodology, which combined many specialised techniques to detect many fault types. A collection of tools was specifically developed to detect cracks, displaced joints, pipe ovality and deposits. Deposits were detected using a collection of Gabor filters, cracks were detected using edged detection and displaced joints were identified by looking for large gaps in the pipe walls. Unfortunately, the work was tested on relatively small datasets (between 10 and 25 images) where the techniques achieved around a 50% accuracy when detecting their respective fault type. Even though these results aren't very impressive, the work does demonstrate the option and merits of combining multiple specialised classifiers into a single streamlined methodology.

2.4 Summary of previous work

Given the variety of previous work in the field of automatic sewer fault analysis, Table 2.1 has been included. This table summarises the key points of all of the previous publications, making it easier to compare and contrast the key details of each piece of work.

***Table 2.1**. Quick reference summary of previous fault detection work, including information on the methodology, testing dataset, accuracy and brief comments on the work.*

Reference	Methodology	Testing Data	Acc. (%)	Comments
(Moselhi & Shebab-Eldeen, 1999)	Applied an edge detector and ANN to analogue images.	A single analogue image containing 8 cracks.	100	Some of the very first fault detection work. Only demonstrated on a single image to detect cracks.
(Moselhi & Shebab-Eldeen, 2000)	Extension using a 3-layer ANN to detect a small range of faults.	218 unseen faults.	98	A marked improvement, showing promise for more than just crack detection.
(Chae & Abraham, 2001)	Applied edge detection to panoramic images recorded by SSET.	172 unseen images.	83	A valid methodology, although it is applied to technology that is not used in common surveying practice.
(Oullette, Browne, & Hirasawa, 2004)	Applied an ANN to raw image segments to detect cracks.	100 unseen image segments containing cracks.	92	Another application of ANNs, however it was only applied to image segments. Making the technique of little use in practice.

Reference	Methodology	Testing Data	Acc. (%)	Comments
(Sinha & Fieguth, 2006)	Applied fuzzy-ANNs to detect cracks.	200 unseen images of a variety of crack types.	95	Good application of fuzzy logic to improve an ANN, although only viable for crack detection.
(Chaki & Chattopadhyay, 2010)	Applied analytical rules to identify cracks using illumination and shape.	5 cracks over 100 contiguous frames.	97%	Although successful, the methodology is limited to well defined cracks and tested on a very small number of examples.
(Khalifa, Aboutabl, & Aziz, 2014)	Applied an analytical approach to detect cracks from detected edges.	101 examples images of cracks.	90%	A direct improvement over Chaki & Chattopadhyay (An intelligent fuzzy multifactor based decision support system for crack detection of underground sewer pipelines, 2010), with improved testing. However, it is still only viable for crack detection.
(Halfawy & Hengmeechai, 2013)	Detected edges and applied a variety of image processing filters to detect cracks.	100 images (mix of normal pipe and cracks).	85	Similar to Khalifa et al. (A new image model for predicting cracks in sewer pipes based on time, 2014) the methodology shows promise for crack detection. However, it was tested using a much more representative dataset including images of normal pipe.
(Jahanshahi & Marsi, 2012)	Constructed 3D scenes of a cracks, using SIFT descriptors. These were classified using a SVM and ANN.	420 synthesized images (50% containing a crack).	78 - 80	A useful technology applied to civil engineering cracks in general. This may be much harder to apply within sewer systems, due to availability of hardware, space within a pipe and the largely uncontrolled nature of the sewer environment.
(Chen, Jahanshahi, & Joffe, 2017)	Detected cracks in metallic cooling pipes, using LBP and a SVM.	~6,000 images (including ~2,000 faults).	76	Again, a useful technology, demonstrated on a substantial dataset. Cooling pipes are a much better controlled environment than sewers, which may make the technology harder to transfer.
(Duran, Althoefer, & Seneviratne, 2003)	Detected faults, using a laser profiler and ANN.	100 readings across a selection of lab-based experiments.	80	An interesting application of new technology. However, laser profilers are still not commonplace in industry and the methodology was only demonstrated in lab experiments.
(Duran, Althoefer, & Seneviratne, 2007)	An improvement on the previous methodology, using multiple ANNs to detect different faults.	Lab studies covering multiple simulated fault types.	80	Additional investigation into the application of laser profilers. With the same shortcomings: experiments on simulated faults

Reference	Methodology	Testing Data	Acc. (%)	Comments
				and reliance on non-standard technology.
(Ahray, Kawamura, & Ishikawa, 2007a)	Compared filtered images of sewer walls to exemplar normal pipe to identify faults.	253 images (147 faults).	84	An interesting and unique methodology. However, it would require a large number of exemplar images to cover the wide variety of shapes/size/materials seen across a network. This could be achieved using a machine learning technique.
(Guo, Soibelman, & Garrett, 2009a)	Applied frame differencing to identify anomalous images.	103 images (57 faults).	84	A great screening technique. Although as it is only anomaly detection no information about a fault is provided and non-faults may be flagged.
(Guo, Soibelman, & J.H., 2009b)	Improved frame differencing, using SIFT features to better compare frames.	192 images (44 faults).	92	An even more effective technique for identifying anomalies, however it will be significantly slower than its predecessor (due to the SIFT calculation), making it less useful for screening.
(Halfawy & Hengmeechai, 2014a)	Identified faults by tracking an operator's camera motion, using optical flow.	350 images.	96	A very unique (and effective) approach to fault detection. However, this relies on the operators motions rather than image contents, making it unviable for full automation.
(Halfawy & Hengmeechai, 2014b)	Identified faults using HOG features and a SVM.	1000 images (~50% contained faults). Only root intrusions were used.	86	A reliable approach for fault detection. Unfortunately, it was only demonstrated on examples of root intrusions.
(Hawari, Alamin, Alkadour, Elmasry, & Zayed, 2018)	Developed an ensemble of specialised detectors. Each one designed for a specific fault type.	10 images of cracks, 32 images of deposits, 22 images of deformation, 32 images of displaced joints.	>50	Although each technique was demonstrated on a relatively small dataset, and achieved below average accuracies, the work did identify the viability of an ensemble of specialised classifiers.

2.5 Image Processing Techniques

2.5.1 Computer vision and object recognition

Many of the most recent developments in the field of automated fault detection have relied heavily on the utility provided by machine learning and computer vision techniques. This interdisciplinary field encompasses a collection of topics identifying the best practice for interpreting digital images and videos. Given the nature of the task, the field of object recognition is most appropriate for the task of automated identification and classification of sewer faults. Currently object recognition has two overarching themes: appearance and feature based techniques. Appearance-based techniques concentrate on the qualities of an image or video, such as frame differencing or gradient matching. Conversely, the more popular feature-based techniques attempt to search for and match key details from one image to another. This is not to say these are the only relevant schools of thought, Deep learning in the form of Convolutional Neural Networks has also proved effective at identifying the contents of a digital image (Krizhevsky, Sutskever, & Hinton, 2012).

Appearance-based detectors have been utilised in previous sewer fault detection methodologies, Guo et al. (2009a) applied frame differencing, whilst Halfawy & Hengmeechai (2013) utilised optical flow. These techniques concentrate on the overall contents of the image, relying on a collection of exemplar images or motion within video footage. For example, frame differencing identifies significant changes between consecutive frames of video footage. This technique is commonly used for motion detection (Lu, Wang, Wu, & Yang, 2008), although the principles have been cleverly reversed by Guo et al. (2009a) for anomaly detection. Alternatively, optical flow identifies the patterns and objects, by the relative motion between frames of video footage. This is applied more widely in the field of object tracking, to identify the position and velocity of an object in video footage. Examples of its application include: recognising human expressions (Yacoob & Davis, 1996) and vehicle tracking (Haag & Nagel, 1999).

More commonly used feature-based techniques rely on the identification of key features within an image, which can be extracted and matched to similar examples. The success of a feature-based technique relies heavily on the correct selection and application of underlying features. These features can be chosen

in many ways and a large collection of algorithms for their generation have been developed, these are more comprehensively discussed in 2.5.2 Feature Descriptors. The second element of a successful feature-based approach is the tool used to match features. This is commonly achieved using a machine learning classifier matching unseen features to those of previously labelled images. However, alternate approaches are equally viable, including the use of clustering or appearance-based paradigms such as frame differencing. Like appearance-based detectors, these techniques have also been applied for the detection of faults within sewer CCTV footage. Halfawy & Hengmeechai (2014b) applied calculated Histogram of Oriented Gradients (HOG) features (Dalal & Triggs, 2005) which were compared to those of labelled images using a Support Vector Machine (SVM) (Cristianini & Sahwe-Taylor, 2000). Guo et al. (2009b) also utilised Scale-Invariant Feature Transform (SIFT) features in combination with frame differencing to produce an improved anomaly detection tool.

2.5.2 Feature Descriptors

As previously mentioned, feature descriptors are the key to many previous successful fault detection and computer vision systems. Feature descriptors enable the quick comparison of digital images in a much lower dimensional space, often simplifying the evaluation process. The broad title of feature descriptors covers a multitude of techniques, each of which aim to represent key details of a digital image. These key details can take many forms, with the application of edge, object, texture, key point and scene detectors being some of the most common.

One of the simplest feature descriptors, and often key in the calculation of other descriptors are edge detectors. These techniques attempt to identify edges in an image, identifying sharp changes in pixel colour or intensity. These discontinuities within an image can be identified using many different mathematical principles, the most common being a gradient based approach. By calculating the pixel gradients across an entire image, edges can be identified using a pre-determined threshold to highlight the steepest intensity gradients. Examples of these include the Sobel method (Gao, Zhang, Yang, & Liu, 2010) or the Canny edge detector (Canny, 1986), both of which have been used to identify cracks within sewer images (Moselhi & Shebab-Eldeen, 1999).

Texture descriptors are another form of feature descriptor, aiming to well represent the different textures or regions within an image. One of the most successful texture descriptors are local binary pattern (LBP) features. LBP works by dividing a given image into a uniform grid of cells and comparing each pixel within the cell to its 8 neighbours, labelling each pixel based on the relative intensity of all comparisons. Once all of the pixel labels within a cell have been identified, they can be counted and formed into a single histogram, defining the contents of a given cell. The final feature vector is then generated by concatenating the histogram of every cell within an image (Oja, Pietikainen, & Maenpaa, 2002). Again, this feature descriptor has been successfully applied to detect the presence of cracks in metallic pipes by Chen et al. (2017).

Given the nature of the work, object detection features, such as the histogram of oriented gradients (HOG) feature has also proved to be effective in the field of computer vision, often complementing LBP features. This technique counts the occurrences of gradient orientation in local sections of an image. It does so by calculating the intensity gradients of all normalised pixels within an image. Like LBP the image is also split using a uniform grid of cells, over which pixels vote on the gradient of the cell. This vote is weighted usually using the magnitude of each pixel's gradient. Next, neighbouring cells are grouped into blocks, with the assigned gradient of its cells being counted to form a histogram of gradients for each block. Finally, all the block's histograms are concatenated to form an overall HOG feature descriptor (Dalal & Triggs, 2005). This object detection-based feature descriptor has been effectively combined with a support vector machine (SVM) by Halfawy & Hengmeechai to achieve an impressive 86% accuracy when identifying root intrusions in CCTV sewer surveys.

The last group of detectors discussed here are commonly known as key point detectors. Scale-invariant feature transform (SIFT) (Lindeberg, 2012) and speeded up robust features (SURF) (Bay, Tuytelaars, & Van Gool, 2006) are the most prevalent of these features. Both SIFT and SURF work by applying identifying maxima and minima, once a collection of gaussian smoothing (Davies, 2004) functions has been applied to the original image. These maxima and minima form the key locations and individual features of the SIFT or SURF descriptor. Each feature is indexed in such a manner that it can be compared across images, enabling similar images to be identified based on the distance

between a feature of the same index. The key difference between SIFT and SURF features is the choice of smoothing operation. SIFT uses cascaded gaussian filters to detect scale-invariant points, whereas SURF applies square-shaped filters as an approximation to gaussian smoothing, making SURF considerably faster. SIFT key point features have been successfully applied by Guo et al. (2009b), matching key points between frames of CCTV footage to identify anomalies within sewer pipes. It is also worth noting that both SIFT and SURF can be used in a similar distributed manner to HOG features, resulting in dense SIFT (D-SIFT) and dense SURF (D-SURF) (Vedaldi & Fulkerson, 2010). Instead of selecting maxima and minima as key points to represent the entire image, D-SIFT and D-SURF utilise a grid of cells. Much like HOG and LBP, a single SIFT or SURF feature is calculated for every cell within an image. This enables D-SIFT and D-SURF to better generalise across images. As the number of features is now pre-defined D-SIFT and D-SURF also often have the additional bonus of being much faster to compute.

2.6 Machine learning techniques

Artificial neural networks (ANNs) are the most common machine learning technique applied across previous work. Inspired by biological brains, this technique learns to perform a given task, training its internal nodal (neuron) structure (Yegnanarayana, 2009). To train a neural network, weights associated with each node are altered so as to match inputs to a specific output. This alteration is the most complex aspect of an ANN application, and is usually achieved using backpropagation (Yegnanarayana, 2009). The neurons within the network are arranged in layers, sandwiched between an input and output layer. A simple ANN contains only a single layer, but more layers can be added. In general, the more layers an ANN contains, the more complex a task it can perform. This behaviour is exemplified by deep learning, where many layers are used to complete extremely complex tasks (LeCun, Bengio, & Hinton, 2015). Examples of deep learning's application range from recognising faces (Parkhi, A., & Zisserman, 2015) to modelling the stock market (Längkvist, Karlsson, & Loutfi, 2014). Although powerful, these techniques are occasionally avoided for their black box nature. The inability to follow an ANN's internal decision-making process makes them unattractive for some applications, particularly where an operator needs 100% confidence in the stability of the technique.

Not as popular, the support vector machine (SVM) is also applied throughout previous work. More specialised than an ANN, the SVM is a supervised classifier, capable of labelling data according to pre-defined categories. To classify data, the SVM compares a datapoint to a learned hyperplane in a higher (possibly infinite) dimensional space. This hyperplane is identified by maximising the margin between classes within a training dataset. Conversely the mapping to a higher dimensional space is achieved using a kernel (Cortes & Vapnik, 1995). Although not as flexible as an ANN, the SVM is much easier to interpret, knowing the kernel and separating hyperplane it is possible to reason through decisions made by the technique, making it more appealing for some applications.

2.7 Contribution

Given the variety of techniques and methodologies previously developed in the area, the work presented in this thesis sets out a new practical solution to the problems with current sewer analysis. The methodologies presented here vary from the majority of previous work in four distinct ways.

Firstly, the work aims to develop a single holistic methodology, capable of performing survey analysis with a single streamlined workflow. This is dissimilar to many other approaches, which tend to be more specialised, detecting a single fault type (often cracks). This specialised approach results in an entire suite of techniques which need to be integrated with each other in order to perform comprehensive analysis (Hawari, Alamin, Alkadour, Elmasry, & Zayed, 2018).

Secondly, the methodologies developed in this thesis are designed to generalise well. Closely related to their holistic nature, the methodologies presented here are developed to work across all types of underground sewer pipe. This includes all combinations of pipe shape, size and material to provide a truly holistic tool. In doing so, a single methodology can be trained on images from one pipe network and applied to multiple other different pipe networks.

Finally, as set out in the aims, the methodologies proposed in this thesis are designed with practicality in mind. The techniques shown here are developed and calibrated using only raw, industry standard CCTV footage. This means that, regardless of the CCTV surveying approach, a company should be able to apply these methodologies to their data, with little to no additional pre-processing. This is dissimilar to previous works which may require the labelling of segmented

images, explicit up-close examples of each fault or additional calibration information for each fault type. Similarly, the tools require a comparatively small calibration dataset, enabling a company of any size to effectively implement the required machine learning elements. Whether the tool works offline or online, it's potential for real time application and ease of calibration are traits rarely explored by previous work.

2.8 Summary

This chapter has discussed previous work in the field of autonomous sewer survey analysis. This led to an overview of the image processing and machine learning techniques which underpin some of the most successful methodologies. A variety of techniques have previously been applied to identify fault in sewer surveys, with a substantial number concentrating on crack detection. Most techniques achieved an accuracy of around 80%, although not all have been tested on a dataset representative of current working practices. These details are all summarised in Table 2.1. As mentioned some of the most successful methodologies extensively applied machine learning and image processing techniques which were also examined. An overview of computer vision was provided, before narrowing down to discuss a multitude of feature descriptors, including LBP, HOG and SIFT. Next the two most dominant machine learning classifiers, artificial neural networks and support vector machines were reviewed. Finally, the novel contributions of the work presented by this thesis were described, identifying the key features which set this work aside from previous studies. The next chapter (Chapter 3: Data) outlines the datasets utilised during the development and testing of this work. This includes the sources of the data, how the data was collected and the process by which example images were extracted and labelled.

Chapter 3: Data

The methodologies established by this thesis are designed to easily integrate with current industry standard surveys. All development and validation has been performed using real CCTV data taken from surveys across the south-west region of the UK. The use of this high-quality data is important given the process's extensive use of the machine learning techniques. To be effective these supervised learning techniques require access to a substantial volume of labelled data similar to that which they are expected to work on.

This chapter describes the data utilised throughout this thesis, briefly discussing the sources, variations in collection techniques and the software currently utilised to manage CCTV surveys. The chapter continues to discuss the process of extraction and labelling, noting both the manual and automated techniques used to generate the core dataset of around 8,000 images. Finally, the core dataset is explored, identifying the wide variety of pipe materials, shapes and sizes, alongside the multitude of fault types and their UK industry standard labels (WRc, 2013).

3.1 CCTV Data Sources

3.1.1 Collection

Currently the collection of most CCTV sewer surveys is performed using remote cameras, which navigate sewer pipes too small for manual human inspection. All current setups employ a digital camera, which records the interior of a pipe, illuminated by an accompanying light. Most of these cameras provide a live feed back to the surveyor, giving a user some control over the camera regardless of the chosen method of navigation.

The simplest method of navigation attaches the camera to a collection of semi-rigid 'push-rods' forcing it through the pipes. Although simple, this method often records the lowest quality footage. Push rods offer little control of the camera's orientation and are unable to navigate hazards or intrusions by any means other than brute force. Furthermore, push-rods have very limited options to be retrofitted with additional technologies, such as laser profilers or recording devices. The more expensive alternative to a push rod system is the use of a pipe

inspection gadget (PIG), not to be confused with those used in pressurised pipes. These devices shown in Figure 3.1 are similar in nature to remote control cars, driving through pipes with a camera and light attached. Unlike push rods, PIGs often enable full control of the camera with inbuilt pan, tilt and zoom functionality. In the hands of a skilled operator the PIG can record high quality footage, providing detailed footage of any suspected fault. Due to their larger nature a PIG also provides a superior platform for additional surveying technologies and can be easily retrofitted to incorporate the most up to date surveying technologies.

Figure 3.1 *An example Pipe Inspection Gadget (PIG). Specifically, this is the IBAK T 76 capable of inspecting pipes with a diameter of 200mm or larger. (IBAK, 2018)*

As important as recording technology is, surveying practices are just as influential on the quality of CCTV surveys. As the UK has many hundreds of thousands of kilometres of pipe only a small fraction of this can be surveyed every year. From conversing with water companies, it is clear to see that this has led to reactive surveying, with most surveys being commissioned to identify a known problem. Similarly, many UK water companies do not have enough internal teams (if any) to perform surveys themselves. Consequently, most surveys are performed by external contractors, who often have differing working practices, even though they all work from the *Manual of Sewer Condition Classification* (WRc, 2013). Typically, survey teams consist of two or three members, with a single person remotely controlling the PIG, and the remaining team members inserting and maintaining the peripheral hardware. Most surveys take place between two

manholes, examining the connecting pipe from upstream. As manholes are not always accessible it is not uncommon for surveys to span multiple pipe sections, passing through intermediate manholes.

Once collected CCTV must be labelled, identifying the type and location of any faults within a pipe. Depending on the method of collection and internal practices of a water company this can be performed in multiple ways, which can be grouped into off and online practices. Offline practices involve the re-watching of an entire CCTV survey, labelling faults after performing a survey. This approach interferes the least with the physical survey and enables a technician to label the footage at their convenience. Alternatively, given the correct setup, a camera operator can label footage as it is collected. As well as being more efficient, it enables a surveyor using a PIG to exploit pan/tilt/zoom features and thoroughly examine a fault before assigning a label. However, the online approach often slows the surveying process, because the camera is usually stationary whilst the operator examines and labels a fault.

3.1.2 The Core (Wessex Water) Dataset

The work presented in this thesis has largely been developed, validated and tested on data provided by the UK water company, Wessex Water. This dataset was provided in multiple blocks over the course of this three-year study. Each block contained a collection of CCTV surveys, some containing only the raw video footage, whilst others included the surveyors' report and accompanying XML files. A total of 613 video sequences was received across all blocks. These provided a representative view of most types of surveys required by the company. These surveys included:

- Regular maintenance surveys, where a team is dispatched to identify a suspected problem or at-risk pipe network. These surveys are the most common and routine for surveyors.
- New pipe surveys, where a section of pipe has just been laid and requires a final inspection before it is put into service.
- Pipe cleaning surveys, where a camera follows closely behind a cleaning device looking for non-serviceable faults. This type of survey was often performed prior to a pipe rehabilitation survey.

- Pipe rehabilitation surveys, where a damaged or degraded pipe has been re-lined and must be inspected for faults or inconsistencies within the new lining.

These surveys were completed by a mix of in house teams and multiple external contractors. Each of which used a different recording setup, resulting in a diverse set of video qualities and styles. Most footage was collected using a PIG, which enabled the use of pan/tilt/zoom functionality, although push rods were still used for pipes less than 150mm in diameter. In both cases footage was labelled during collection, with the camera pausing to allow a surveyor to record observations. All survey labels were assigned according to the same standard outlined in the *WRc Manual of Sewer Condition Classification* (WRc plc, 2013). However, due to the nature of these labels and the differing levels of experience across operators, annotations are often subjective.

Due to the large number of different pipes and faults contained within this dataset, it is ideal for the development of a generalised fault detection method. This ensures that the technology can be demonstrated over many of the different use cases of sewer surveys in practice. On the other hand, due to the diluted nature of the dataset, there may be an insufficient number of examples of some pipes/faults to comprehensively train some machine learning algorithms. This could lead to reduced performance when identifying poorly represented faults or other less common observations. Details concerning the distribution of faults in the Wessex Water dataset can be found in section 3.2.

3.1.3 Additional Datasets

Over the course of the project, multiple other companies (from across the globe) kindly provided data to validate the developed techniques. From the UK this included the University of Exeter and the water company, South West Water. Exeter University provided CCTV surveys of their Streatham campus sewer system. These routine surveys must be performed every few years and were collected by Exjet, a local surveying contractor. Due to the era in which the university's sewers were constructed, many of the pipes are made from pitch fibre. This material was briefly used between 1960 and 1970, before being replaced by modern plastics. This makes pitch fibre extremely rare in public sewer pipes, relegating this dataset to edge case testing. Conversely, South West

Water provided a small collection of surveys on more standard sewer pipes. These surveys were performed by external contractors and analysed internally. As such, these videos provided an excellent external dataset for testing the methodology's performance. From outside the UK, small datasets were provided by HSY (Helsinki region environmental services authority) in Finland and City West Water in Melbourne, Australia. Both of these datasets provided good quality footage on standard CCTV pipes. These surveys provided a good dataset for the external testing of the techniques developed in this thesis. These additional datasets were used throughout the development of the methods presented in this thesis, ensuring each generalised well across different sewer networks.

3.1.4 Format

Across the UK and much of the world, sewer CCTV surveys are produced in uniform format. This always includes the raw CCTV video (.mpg or .mp4) inspection and an accompanying surveyors' report (.pdf). It should be noted that due to the compression used in .mpg and .mp4 formats, the image quality is lower than what would be expected from a normal .jpg image, which is itself compressed and may lead to "blocky" compression artefacts. Specialised annotation programs are used to record and generate sewer survey footage and annotations, most commonly WinCan (CD Lab AG, 2018). This software enables the annotation of CCTV footage on and offline, automatically generating required documentation, including the surveyors' report. Due to the internal processing of the WinCan software, additional XML files are built, providing details about each fault in an easy to interpret format. XML, which stands for extensible mark-up language, allows a user to define a format for storing data, ideal in the case of the fault definitions given by WRc (2013)

3.2 Extraction and Labelling

Given that over 50 hours of CCTV survey footage in the Wessex Water dataset, a dataset of example images was extracted and labelled. This provided the foundation for the development of all fault analysis methodologies covered by this thesis. Continuous CCTV footage is usually recorded at 25 fps (frames per second), implying the raw footage provided by Wessex Water contained over 4.5 million images. To make this of any use to the fault analysis techniques was reduced to a much smaller number of labelled exemplar images.

3.2.1 Still images

To create a dataset of labelled still images, a combination of automatic and manual frame extraction was performed. Once extracted, frames were manually labelled according to the WRc guidelines (WRc, 2013). Automatic extraction was performed, where available, using the XML files accompanying some surveys. By automatically reading these files, the timestamps of faults within the CCTV videos could be identified. Due to the coarse time resolution and continuous nature of faults, this timestamp often spanned multiple frames. In these cases, the middle frame of a fault's duration was extracted. Once complete, all extracted frames were manually reviewed, discarding poor quality frames. A frame was defined as poor quality if its contents were visually unclear, contained no fault or were unidentifiable by a human. As not all videos were provided with accompanying XML files, images containing faults were also manually extracted. This was achieved by reviewing every 25th frame or one frame every second. At this stage, high quality images containing faults and images of normal pipe were selected for use in the final dataset. Using both the automated and manual techniques 7,953 images were extracted of which 2,424 images contained faults.

Once extracted, all frames required labelling. This label provided information on a frame's contents, including whether a frame contained a fault, and if so the fault type. The *Manual of Sewer Condition Classification* (WRc, 2013) contains definitions for many different types of fault and many more subcategories for each. Consequently, only the core fault category was labelled for each fault, these and the subcategories covered by the labelling are provided in Table 3.1. Example images of each fault type, taken directly from the Wessex Water dataset, are presented in Figure 3.2. Faults within pipes often appear together, whether they are related or not; many images contained examples of multiple faults. These were labelled with each fault being individually identified, as a surveyor would.

Table 3.1 *Breakdown of the distribution of fault types in the Wessex Water dataset. Note that distribution (%) totals more than 100% as multiple fault types may appear in a single image.*

Fault Type	Sub Categories	Number of occurrences	Distribution (%)
Joint	Displaced, Open	924	38.1
Deposits	Attached, Settled	454	18.7
Crack / Fracture	Longitudinal, Circumferential, Multiple, Radiating, Spiral	322	13.3
Surface	Spalling, Corrosion, Visible aggregate/reinforcement, Blister/Bulge	293	12.1
Roots	Fine, Tap, Mass	274	11.3
Infiltration	Seeping, Dripping, Running, Gushing	142	5.9
Obstruction	Intruding junctions, Masonry, Protrusions	98	4.0
Broken / Collapsed	-	61	2.5
Other	Vermin, Lining, Defective Repairs	49	2.0
Hole	-	37	1.5
Brickwork	Missing mortar, Displaced bricks, Missing bricks	22	0.9
Deformation	Vertical, Horizontal	9	0.4

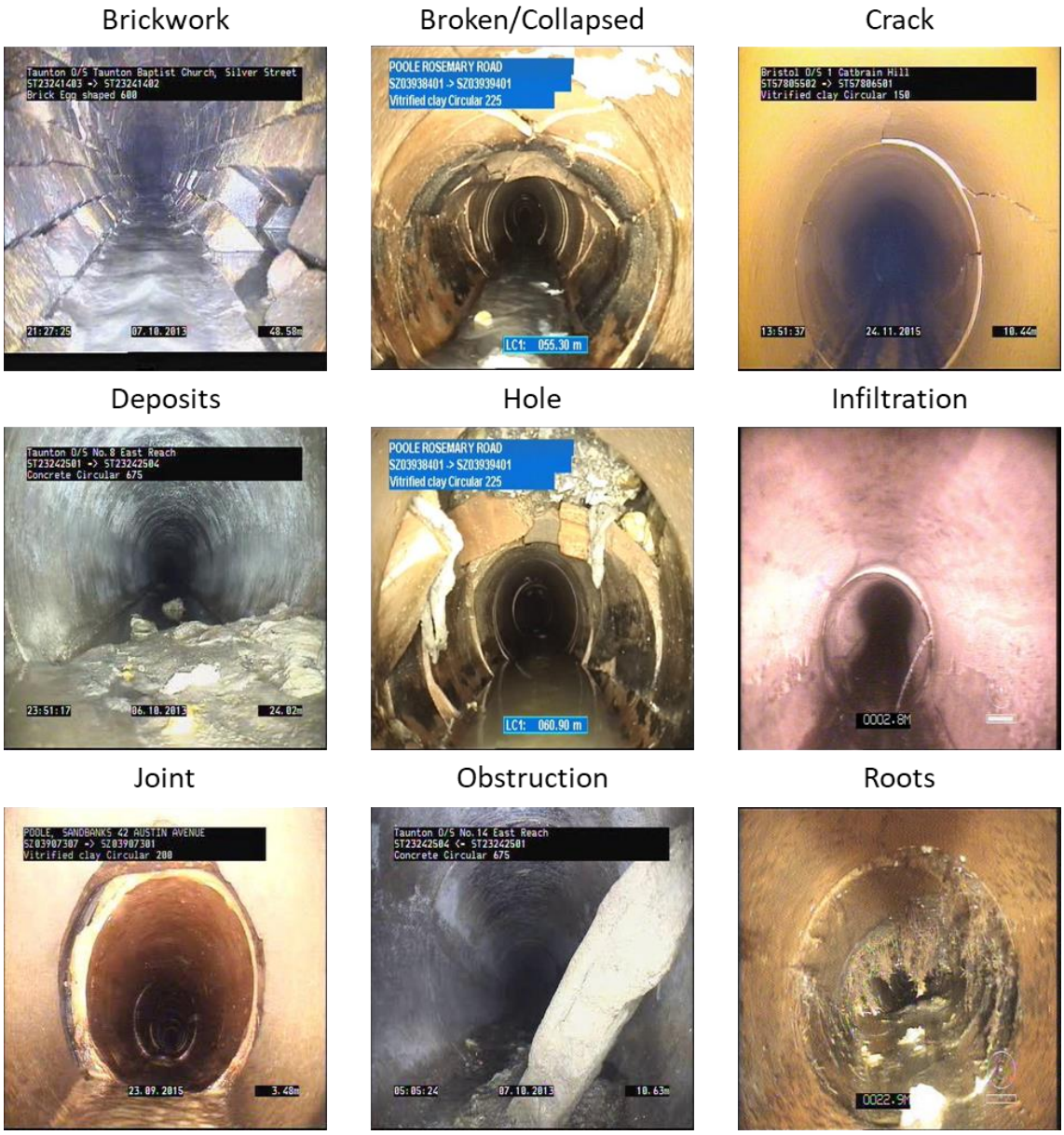


Figure 3.2 *A collection of example images containing various fault types, extracted from the Wessex Water dataset.*

In addition to the large variety in fault types, the datasets provided a diverse selection of pipe size, shape and material. Pipe diameters ranged from 150mm pipes, used in domestic sewers, to large 1,500mm sewer mains. Pipe materials included: vitrified clay, concrete, cast iron, brick and PVC. Pipe shapes also varied, ranging from standard circular pipes, to horseshoe and egg-shaped pipes. Conversations with Wessex Water revealed that these fault type, size, material and shape distributions are representative of their diverse network.

3.2.2 Video sequences

Throughout the thesis, the fault analysis methodologies have been tested on continuous CCTV footage. This was done to evaluate their effectiveness when applied as the technologies would be in practice. To do so a three CCTV videos were also labelled. This labelling was done at a resolution of 1 second (25 frames) and contained all the details present in the still image labels. In order to ensure training and validation sets were not mixed, around 5% of videos were set aside solely for testing from the 613 videos in the Wessex Water dataset. In addition, the extra videos provided by the University of Exeter, South West Water, HSY and City West Water were also kept separate, for validation only.

3.3 Summary

To summarise, this chapter briefly details the datasets used to develop and evaluate the work presented by this thesis, alongside their methods of collection. Industry standard practices and technologies have been overviewed, identifying the key differences between online and offline annotation. Each of the sources of the data has been discussed, with specific emphasis on the large dataset, containing over 50 hours of survey footage, provided by Wessex Water. The process of extracting and labelling examples from raw CCTV footage has been discussed for both still images and continuous example footage. Finally, the 8,000 images of the core dataset have been presented, identifying the many fault types and their distributions. The next chapter (Chapter 4: Fault Detection), puts this dataset to work, defining and evaluating the fault detection methodology in its simplest form.

Chapter 4: Fault detection

4.1 Introduction

An engineer learning to analyse CCTV surveys in the UK must complete a week-long training course. This course teaches an engineer how to identify a fault, distinguishing between fault types and sub-categories, alongside learning best surveying practices. Unfortunately, this process is much harder to implement digitally. Computers must learn to interpret complex images before any analysis can occur, something humans learn from the day they are born. In order to simplify the automated CCTV analysis process, the procedure was broken down into two simpler stages: fault detection and fault classification. The first fault detection stage is described here.

The fault detection methodology aims to identify the presence of a fault in a single image, which can in turn be applied to continuous CCTV data. The method was developed, keeping the key principles and aims defined in Section 1.4.2. These principles include the development of a holistic tool, capable of generalising to all pipe shapes, materials and sizes. Similarly, the method should be flexible enough to integrate with any company's current working practices. As such, the tool can perform on or offline, identifying faults in footage as it is recorded or after collection.

The method itself, utilises the principles of computer vision as outlined in Section 2.5. More specifically, it applies a machine learning classifier to feature descriptors calculated from CCTV images. Similar in nature to the works of Halfawy & Hengmeechai (2014a) and Chen et al. (2017), this combination has performed well, achieving accuracies of at least 80%.

4.2 Fault Detection Methodology

4.2.1 Overview

The fault detection methodology applies modern image processing and machine learning techniques to raw CCTV footage taken directly from industry surveys. The method is data-driven in the sense that relevant image characteristics and classifier parameters are learned from a database of images, which have been labelled as containing a fault or normal pipe. It is therefore assumed that such a

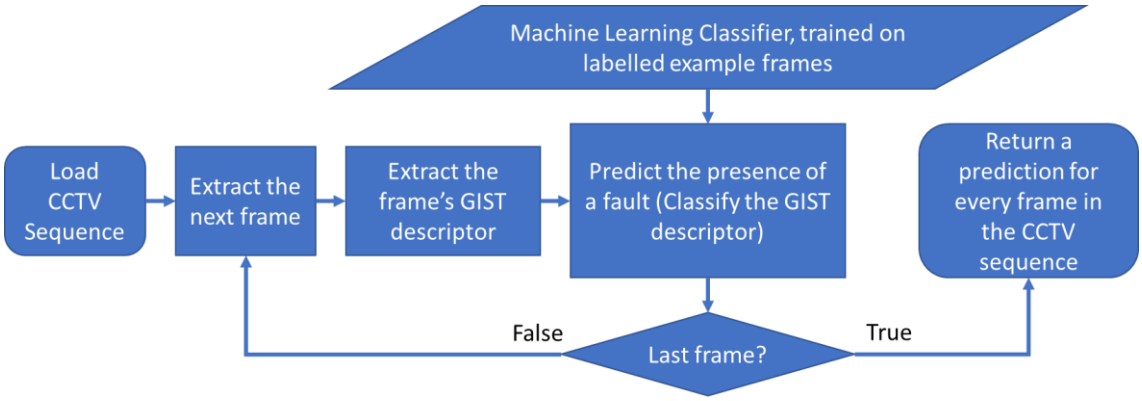


Figure 4.1 *Flow chart defining the structure of the fault detection methodology, and its online implementation.*

database is available for the effective implementation of the tool. In practice, this database will constantly grow, expanding as new faults are identified and verified by a technician. These extra samples should lead to an improved detection accuracy when the methodology is retrained on its corrected misclassifications. The training process, although relatively time consuming (less than an hour) to complete, does not require human intervention and so could be performed periodically overnight.

As illustrated in Figure 4.1, the online method examines each frame, predicting the presence of a fault in each frame in turn. This structure can be intuitively broken down into three stages: 'Frame Extraction & Pre-processing', 'Feature Extraction' and 'Classification' each of which is described below.

4.2.2 Frame extraction and pre-processing

In order to process frames individually, each frame is extracted from the raw CCTV footage as an RGB image. Surveys are generally recorded at a nominal rate of 25 frames per second (fps). Since, as shown later in Section 4.4.2, the computational effort to process each frame is light, all frames are extracted. While it would be possible to omit some frames, excessive reduction of the sampling frequency (F) will detrimentally impact performance. In the work presented here, all recorded frames were used, as shown in Table 4.1.

Each extracted frame is converted to greyscale and resized to a lower resolution (R). Experimentation showed that retaining colour information degrades classification performance, likely because of the wide variation in illumination encountered in real surveys. Similarly, resolution is reduced as experiments

showed that higher resolution images provided little performance improvement, whilst substantially increasing the processing time.

4.2.3 Feature Extraction

The visual content of each frame is characterised using GIST feature descriptors (Oliva & Torralba, 2001), named for their ability to represent the overall gist (i.e. the essence or substance) of an image. This feature descriptor represents the composition of the image in a much lower dimensional space. A standard GIST descriptor represents the frame as a 512-dimensional vector (c.f. the 128 x 128 ~ 1.6×10^4 pixels comprising even the reduced resolution image). Other common feature descriptors, such as HOG (Dalal & Triggs, 2005) and SIFT (Lindeberg, 2012), focus on local image patches and therefore yield a much higher dimensional image descriptor. Preliminary experiments showed that HOG and SIFT descriptors, at best, performed similarly to GIST, whilst taking much longer to process. Some of the GIST descriptor's success could be attributed to its holistic nature, aiming to describe the overall composition of the image. This defining attribute is clear, given the GIST descriptors success in various applications, including development of self-driving cars (Pugeault & Bowden, 2011) to the classification of panoramic landscapes (Murillo, 2013).

The GIST feature descriptor describes the energy of square patches of the image at a range of orientations and scales. To achieve this, the greyscale frame is convolved with a series of Gabor filters (Bovik, Marianna, & Wilson, 1990). Following Oliva & Torralba's (2001) example, these filters are arranged at 4 scales (S) and 8 orientations (θ) (Figure 4.2), resulting in a series of 32 feature maps characterising the oriented-energy at the 4 scales. The GIST feature descriptor is then formed by the squared response to each filter summed over each cell in a 4x4 grid covering the image (Figure 4.2), thus forming a 512-dimensional image descriptor. Experimentation showed that an increased number of cells, frequencies or scales did not improve classification accuracy.

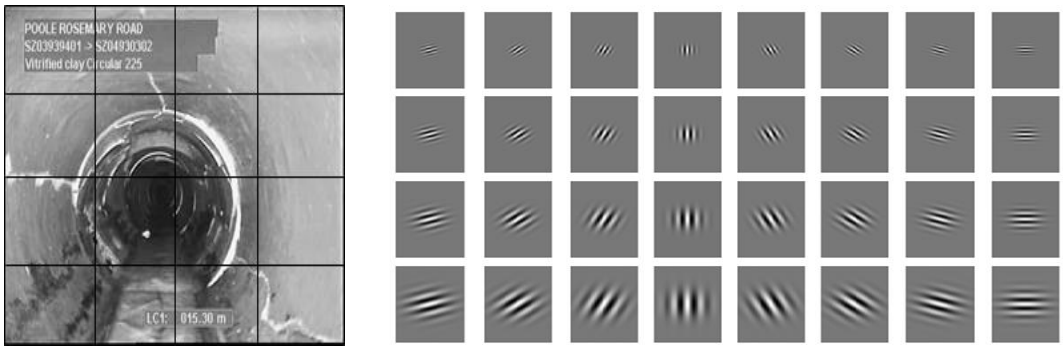


Figure 4.2 *(Left) Greyscale frame, overlaid with the 4x4 grid of square cells. (Right) Gabor filters at 4 scales (S) and 8 orientations (θ).*

4.2.4 Classification

Having calculated a frame's GIST descriptor, a prediction about the frame's contents can be made by a trained machine learning classifier. This classifier estimates the probability of a frame containing a fault. The work presented here uses a Support Vector Machine (SVM) or a Random Forest (RF), both of which are trained on a dataset of previously labelled frames.

The SVM was chosen for its robust nature, and prevalence in the literature, being applied in some of the most recent and successful methodologies (Jahanshahi & Marsi, 2012) (Halfawy & Hengmeechai, 2014b). The SVM predicts the class of a frame (i.e. containing a fault or not) using a learned hyperplane in a higher (possibly infinite) dimensional space (Cortes & Vapnik, 1995). In this higher-dimensional space the SVM is trained using a labelled dataset (see Chapter 3) by maximising the margin between the defined classes (Hearst, Dumais, Osuna, & Platt, 1998) (Steinwart & Christmann, 2008). Calculations in the high dimensional space are efficiently carried out in the original low-dimensional space via a kernel. In this work the radial basis function (RBF) kernel was used, although other nonlinear kernels (e.g., polynomial, sigmoidal) yield comparable results. The choice of kernel and other internal parameters such a C and γ were automatically learnt using cross validation (Kohavi, 1995). The trained classifier is then used to predict the class of unseen frames according to its position relative to the separating hyperplane (Cristianini & Shawe-Taylor, 2000). In addition, the estimated probability of a frame belonging to the 'faulty' class is generated based on its distance to the separating hyperplane; this was done using Platt's method (Platt, 2001).

On the other hand, a Random Forest (RF) is an ensemble classifier, utilizing a collection of decision trees to predict the class of unseen frames (Breiman, 2001). After comparing multiple RF architectures and training methods, the Extra Trees algorithm (Geurts, Ernst, & Wehenkel, 2006) was selected for its superior performance. This technique utilises the same dataset of labelled frames as the SVM to train a collection of decision trees. Each tree in the decision forest is trained individually on a randomly-chosen subset of the 512 GIST features; moreover, decision thresholds in the tree are selected from a randomly generated set of thresholds so as to maximise the normalised information gain. After every tree in the forest has been trained, unseen frames can be classified: the GIST descriptor of the frame is processed by each individual tree, after which the forest votes on that frame's class; the proportion of votes in favour of a class can be interpreted as an estimated probability of class membership. Geurts et al. (2006), show that this strategy results in a random forest that classifies accurately and generalises well to new data. Finally, the RF's structure naturally works well as a multi-class classifier. This is dissimilar to the SVM which performs binary classifications, and is an extremely useful feature, utilised later in the fault classification method (Chapter 6).

4.2.5 Parameters

Table 4.1 *A description of fault detection methodology parameters and default values used in below case studies and justification for these chosen values.*

Variable (value used)	Description and Justification
F (25)	The rate at which frames are extracted from the raw CCTV footage (fps). When applied to continuous CCTV footage, every recorded frame was processed. This is done as the methodology proved to be fast enough to be ran in real-time, giving little benefit to skipping frames.
R (128x128)	The resized resolution of the extracted frame. As most footage provided by Wessex Water was recorded in a minimum resolution of 512 x 512, the images are reduced to a quarter of the minimum resolution: higher resolutions showed little improvement in performance, whilst dramatically increasing processing times.
Θ (8)	The number of orientations of Gabor filter used to calculate the GIST feature descriptor.

	Following the parameters of the GIST feature descriptor used by Oliva & Torralba (2001), eight orientations of Gabor wavelet were used. As the wavelets are symmetrical, each wavelet was a rotation of the previous by 22.5˚, as shown in Figure 2.
S (4)	The number of scales of Gabor filter used to calculate the GIST feature descriptor. Again, following the example of Oliva and Torralba (2001), four scales of Gabor wavelet were used.
Cells (4x4)	The resolution of the grid used to calculate the GIST feature descriptor. A 4x4 grid is used as specified by Olive & Torralba (2001). Higher resolutions were tested, producing little improvement in performance whilst dramatically increasing processing times.

4.3 Case Study: Still CCTV images

This first case study examines the application of the new fault detection methodology in its rawest form to example CCTV frames. These frames are still images, previously extracted from standard CCTV footage, as explained in section 3.2. The case study demonstrates the application of both the Random Forest (RF) and Support Vector Machine (SVM) classifiers, comparing and contrasting their effectiveness.

4.3.1 Data and configuration

The fault detection methodology was calibrated and tested using the dataset of 7,953 still images detailed in Section 3.2.1. Excluding the parameter F, which is not relevant as images have already been extracted, the default parameters given in Table 4.1 were used for this case study. Similarly, the RF contained 100 trees (see Figure 4.3) and the SVM used a radial basis function (RBF) kernel with a regularisation constant C of 150 and γ of 0.002. These parameters were learnt using a system of cross validation. In order to make the most of this dataset, 25-fold cross validation was used to separate the frames into training and testing sets (Kohavi, 1995). Cross validation split the randomly shuffled dataset into 25 equally sized groups (i.e. folds). Each of the 25 folds was in turn set aside to form the test dataset, whilst the remaining 24 folds were used to train the classifier.

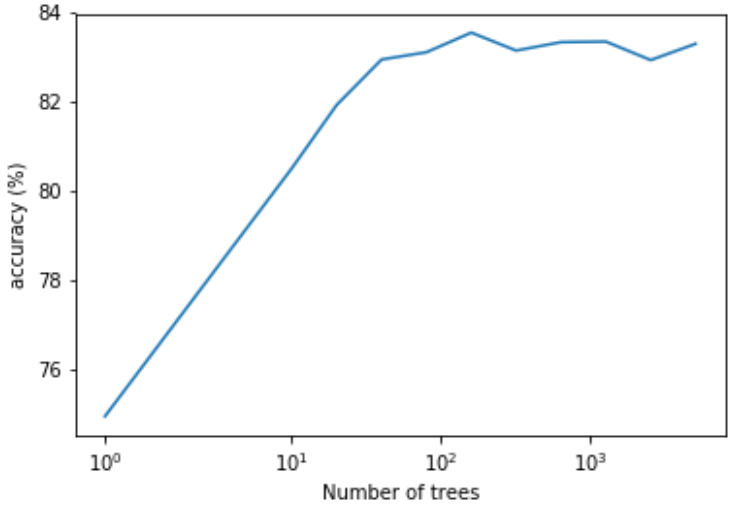


Figure 4.3 *Plot comparing the number of trees in a Random Forest, vs the accuracy achieved in preliminary fault detection experiments.*

The generalisation accuracy of the method was then estimated by averaging the accuracy over all 25 validation sets. Accuracy was defined as the percentage of correctly classified (i.e. detected) faults, using each frame's label as the ground truth. When evaluating each approach, speed and accuracy were both considered. Since the methodology estimated a probability of a fault's presence, frames are considered faulty if this probability exceeded a threshold value of 0.5. However, the performance over the full range of thresholds was examined using Receiver Operating Characteristic (ROC) curves (Fawcett, 2006).

4.3.2 Results and Discussion

Independently applying the SVM and RF classifiers, within the fault detection methodology, each achieved accuracies of 78.7% and 83.0% respectively. These high obtained accuracies indicate the method is sufficiently accurate to assist surveyors in the location of faults (of all types) within sewers. Comparing these accuracies to other similar works, Chen et al. (2017) achieved an accuracy of around 76% using a similar dataset of ~6,000 images (including ~2,000 faults). Similarly, Halfawy & Hengmeechai (2014b) achieved an accuracy of 86% on a smaller dataset of 1,000 images (including 500 faults), which only contained root intrusions. Furthermore, preliminary experiments applying a convolutional neural network (CNN) to the same task, yielded an accuracy of 81% (see section 5.6.3). Widely considered one of the most powerful image processing techniques, CNNs are a branch of deep learning specialised to the analysis of visual imagery

(Krizhevsky, Sutskever, & Hinton, 2012). In this application, a MobileNet (Howard, et al., 2017) CNN's final layer was re-trained and tested on sewer CCTV images in the same manner as the SVM and RF. Continuing to evaluate the performance of the SVM and RF on the basis of fault type, it is clear to see that some of the poorly represented faults (<50 examples) are harder to detect. However, other better represented fault types were also misclassified, including 'Joint' and 'Crack' faults.

Table 4.2 *A breakdown of the SVM and RF classifiers misclassifications. 'Total' refers to the total number of failures of a given fault type, this is accompanied by the rate at which a fault type was misclassified. 'Dataset Total' refers to the total number of occurrences of a given fault type within the entire dataset. It should also be noted that the 'None' fault type refers to normal pipe, whilst the 'Multi' fault type refers to an image containing many different faults.*

Fault	SVM		RF		Dataset Total
	Total	(%)	Total	(%)	
None	315	(5.7)	417	(7.5)	5529
Multi	144	(46.8)	87	(28.2)	308
Deposits	209	(54.9)	146	(38.3)	381
Surface	140	(58.3)	10	(4.2)	240
Joints	505	(66.2)	344	(45.1)	763
Obstructions	32	(47.1)	24	(2.8)	68
Crack	160	(1.5)	100	(41.3)	242
Roots	90	(45.2)	54	(37.5)	199
Infiltration	44	(37.9)	34	(29.3)	116
Hole	11	(50.0)	12	(54.5)	22
Deformation	4	(57.1)	3	(42.9)	7
Broken	12	(37.5)	3	(9.4)	32
Brickwork	10	(83.3)	8	(66.7)	12
Other	18	(52.9)	11	(32.4)	34

Given the limited number of examples, it is expected that 'Hole', 'Deformation', 'Broken', 'Brickwork' and 'Other' faults are commonly misclassified. This would be because both the SVM and RF are data driven techniques, requiring a sufficient number of labelled examples to perform well. Outside of these, 'Joint'

faults were the most misclassified with rates of 66.2% and 45.1% for SVM and RF respectively. As this is the best represented class of fault, it is likely that these misclassifications are due to the fault's subtle nature as demonstrated by Figure 4.4. This could also be the case for the 'Crack' fault type too, given the plethora of literature detailing crack detection (Section 2.2).

In order to detect more of these subtle faults, the probability threshold used to define the presence of a fault could be reduced, identifying more faults, at the cost of misclassifying more normal pipe (See Figure 4.5). Alternatively, a specialised detector could be applied alongside the methodology, such as Chen et al.'s (2017) crack detector, although this would dramatically reduce the computational efficiency of the methodology, eliminating any chance of real-time application.

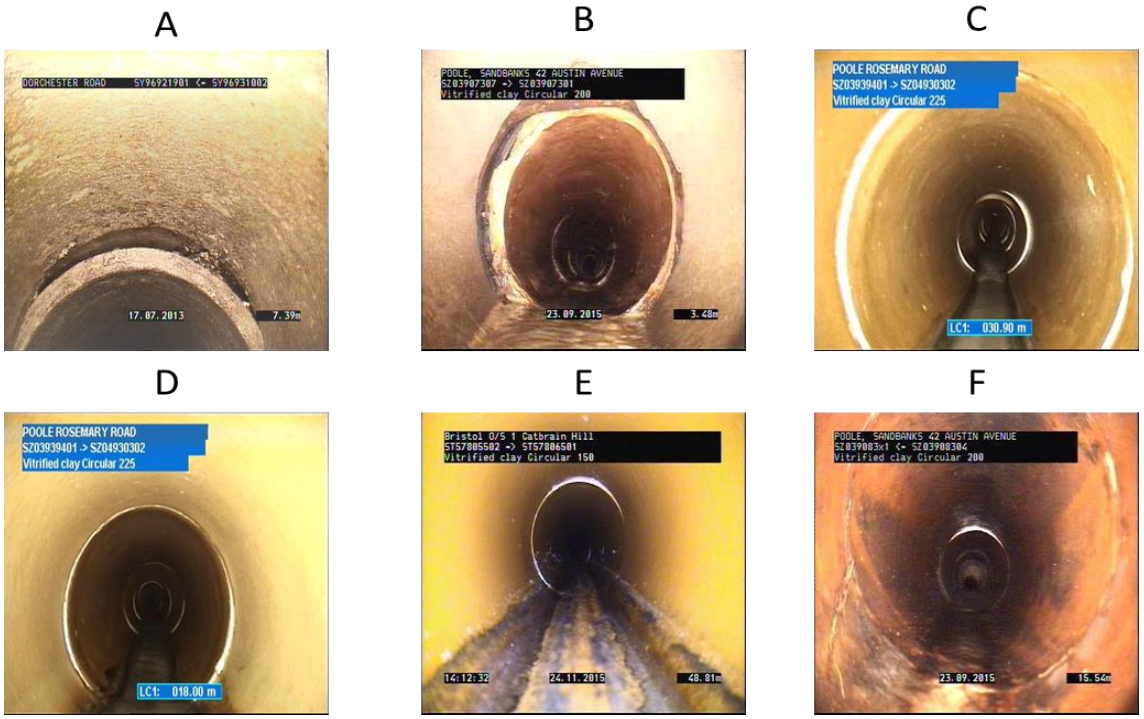


Figure 4.4 *A comparison of 'Joint' faults, highlighting the subtlety in their appearance. Images A and B are the most obvious 'Joint' faults and were correctly identified. However, images C, D and E are labelled (by the operator) as displaced joints, even though they are almost indistinguishable from a normal pipe joint. As such they were misclassified by the methodology. Finally, Image E contains not only a displaced joint, but also a root intrusion. Due to the texture and colour of the pipe wall this would be hard to notice for a human being, let alone the methodology, as such this image was also misclassified. Even though faults C, D, E and F are misclassified all of the faults are low in severity and would not require intervention.*

Again, comparing the performance of the two classifiers it is clear to see the RF usually performed the best, only outperformed by the SVM when identifying normal pipe and 'Hole' faults. Even then, the SVM only achieved a 5.7% misclassification rate on normal pipe as opposed to the RF's 7.5%. Conversely, the RF achieved a substantially lower misclassification rate for most fault types. Most impressively the RF achieved a very low misclassification rate of 4.2% on 'Surface' faults, unlike the SVM's misclassification rate of 58.3%. These differences are made even clearer when examining the overall misclassification rates of each technique.

Tables 4.3 and 4.4 show the confusion rate matrices for the SVM and RF classifiers. The False Positive Rate (FPR) is the rate at which a fault is predicted when no fault is observed (upper right entry in the matrix) whilst the False Negative Rate (FNR) is the rate at which no fault is predicted when a fault is actually present (lower left entry in the matrix). As it can be seen from Tables 4.3 and 4.4, the FPR for the SVM is 21%, marginally lower than that of the RF's FPR of 23%. On the other hand, the RF achieves a substantially lower FNR, 16% to the SVM's 26%. This implies the RF is much better at identifying faults within a frame, and as such is less likely to miss a fault when applied in practice. These results correlate well with the results found in the breakdown of misclassifications in Table 4.2.

Table 4.3 *Confusion rate matrix for the fault detection methodology, using the Support Vector Machine (SVM) classifier, when applied to the core Wessex Water dataset.*

		Actual	
		Fault	Normal
Prediction	Fault	0.74	0.21
	Normal	0.26	0.79

Table 4.4 *Confusion rate matrix for the fault detection methodology, using the Random Forest (RF) classifier, when applied to the core Wessex Water dataset.*

		Actual	
		Fault	Normal
Prediction	Fault	0.84	0.23
	Normal	0.16	0.77

Use of the accuracy as a measure of a classifier's performance is tantamount to assuming that the costs of misclassification (false positives and false negatives) are equal. However, as misclassification costs are seldom equal, and in light of the fact that in continuous CCTV footage the proportion of normal frames is likely to far exceed the proportion of faulty frames, the performance of the classifiers is displayed as Receiver Operating Characteristic (ROC) curves (Fawcett, 2006) in Figure 4.5. The ROC curves are constructed by varying the decision threshold from small (so that almost all frames are classified as faulty corresponding to a high false negative cost) to large (so that almost all frames are classified as normal, corresponding to a high false positive cost). An ideal ROC curve would push towards the top left of the graph, achieving a True Positive Rate (TPR; rate at which a fault is predicted when fault is observed) of 1.0 for a FPR of 0.0. Both techniques push towards this goal, although it is clear that the RF is superior, as its curve dominates the SVM's at all threshold values. This is confirmed by the difference in the areas under the curve (AUC), which summarises the classifier's performance over all misclassification costs. The RF achieves an AUC of 0.89, whilst the SVM achieves only 0.81. This domination of the SVM by the RF clearly shows the superiority of the RF for the task of fault detection.

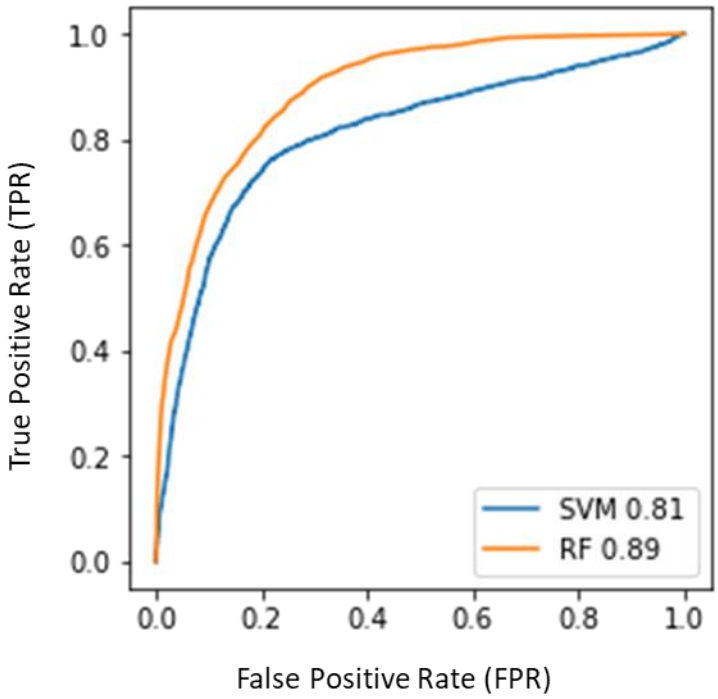


Figure 4.5 *Receiver Operating Characteristic (ROC) curves for both the Support Vector Machine (SVM) and Random Forest (RF) classifiers, when detecting faults in the core Wessex Water dataset.*

Finally, considering the computational efficiency of each technique (excluding the calculation of GIST descriptors), the SVM processed around 160 frames per second (fps), whereas the RF processed around 35 fps. These results were achieved on a standard laptop computer (i5 2.3 GHz processor and 8GB RAM), using a single core. Given that CCTV footage is typically recorded at 25 fps, both of these techniques should be capable of processing footage in real-time as it is recorded, enabling online CCTV analysis. Even if these calculations do not consider the additional delay incurred during the I/O process of recording footage, the SVM should still be able to achieve real-time processing, whilst the RF could be easily optimised to achieve much faster classifications and it is eminently suitable for parallelisation. Furthermore, a recording rate of 25 fps may be unnecessarily fast given the slow pace at which the camera travels. This pace means faults are typically visible across many frames, potentially allowing for a rate of frame extraction (*F*).

4.3.3 Summary

Overall the fault detection method has been successfully applied to still CCTV images. Furthermore, both the RF and SVM have proven to be effective classifiers, capable of detecting sewer faults. Each technique achieved accuracies comparable to those found in the literature, with the SVM and RF achieving accuracies of 73.7% and 83.0% respectively. Both techniques have proven to be sufficiently fast for real-time processing of CCTV, even though the SVM was significantly faster. These results indicate the potential of the system, as such the fault detection method is demonstrated on continuous CCTV footage in a second case study (section 4.4).

4.4 Case Study: Continuous CCTV footage

This second case study aims to demonstrate the application of the new fault detection methodology on continuous CCTV footage. The method is applied to the three fully labelled Wessex Water videos, unused in the creation of the dataset of still images. As with the first case study (section 4.3) both the Random Forest (RF) and Support Vector Machine techniques are evaluated, so as to compare their effectiveness when applied as they would be in practice.

4.4.1 Data and Configuration

The fault detection methodology was calibrated using the entirety of the core Wessex Water dataset (Section 3.2.1), requiring no cross validation. The three surveys used in this study vary in length and fault types, which are typical of most sewer surveys in the UK (see table 4.5). For comparison the surveys have been labelled using the surveyors' annotations, at a temporal resolution of 1 second. Although used as the ground truth in this case study, surveyors' annotations are often imperfect and not all faults are labelled by the surveyor, as shown in Figure 4.6.

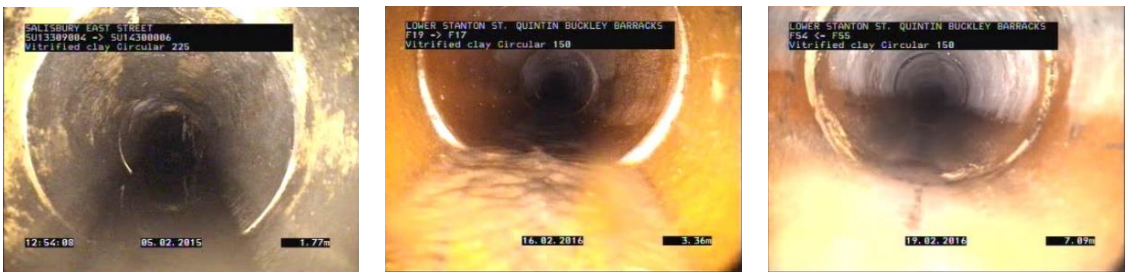


Figure 4.6 *Three subtle faults (one from each video, unlabelled by the surveyor. This mislabelling often only occurs with very minor 'joint' or 'deposit' faults.*

The methodology was applied using the default parameters defined in Table 4.1. Similarly, the RF contained 100 trees and the SVM used a radial basis function (RBF) kernel with a regularisation constant, *C,* of 150 and γ of 0.002. As with the previous case study, the accuracy was used to evaluate its effectiveness. Given the methodology estimated a probability of a fault's presence, frames are considered faulty if this probability exceeds a threshold of 0.5. As before the performance over the full range of thresholds was also examined using Receiver Operating Characteristic (ROC) curves.

Table 4.5 *Table summarising the contents of each of the three testing videos. Number of faults indicates the total number of blocks of fault labels throughout the entire video sequence.*

Video Number	Duration (min:sec)	Pipe Material	Number of Faults	Example Frames
1	2:11	Vitrified Clay	6	 
2	1:59	Vitrified Clay	12	 
3	6:43	Vitrified Clay	29	 

4.4.2 Results and Discussion

To identify the effectiveness of the methodology and each of the classification techniques, the frame by frame accuracy was calculated over each of the three

surveys. The RF performed the best on videos 1 and 3, achieving accuracies of 84.7% and 83.6%. On the other hand, the SVM performed substantially worse on video 1, with an accuracy of 69.8% and marginally worse on video 3, with an accuracy of 79.0%. Generally speaking, these results are in-line with those achieved when applied to still CCTV images, demonstrating the stability of the tool.

However, both techniques performed much worse than anticipated when applied to video 2. The RF achieved an accuracy of 53.0%, whilst the SVM achieved a much lower accuracy of 43.5%. When examining video 2, it appears that many of the faults are subtle 'Joint' faults. As such it is unsurprising that the methodology performed poorly in this video as it also performed poorly when identifying similar faults in still images (see Section 4.3.2). Fortunately, many of these faults are of low severity (grade 3 or less) and would not require immediate maintenance (see Figure 4.4).

Digging deeper into the misclassifications of each tool, confusion rate matrices (Tables 4.6 – 4.11) and ROC curves (Figures 4.4 – 4.6) have been calculated for each technique applied to each video. Examining the confusion matrices for video 1, show that each classifier achieved the same FPR of 2%, correctly classifying normal pipe 98% of the time. However, the RF achieves a FNR of 49%, lower than the SVM's 66%. Although neither of these rates are low, both techniques appear to perform reasonably well, this is confirmed by the ROC curves (Figure 4.7). Although the RF's curve does not entirely dominate that of the SVM, it is preferable for most acceptable false positive rates (FPR). This is further reinforced by the RF's marginally higher AUC (area under the curve) of 93% as opposed to the SVM's 90%.

Table 4.6 *Confusion rate matrix for the fault detection methodology, utilising a Support Vector Machine (SVM) classifier, applied to video 1.*

		Actual	
		Fault	Normal
Prediction	Fault	0.34	0.02
	Normal	0.66	0.98

Table 4.7 *Confusion rate matrix for the fault detection methodology, utilising a Random Forest (RF) classifier, applied to video 1.*

		Actual	
		Fault	Normal
Prediction	Fault	0.51	0.02
	Normal	0.49	0.98

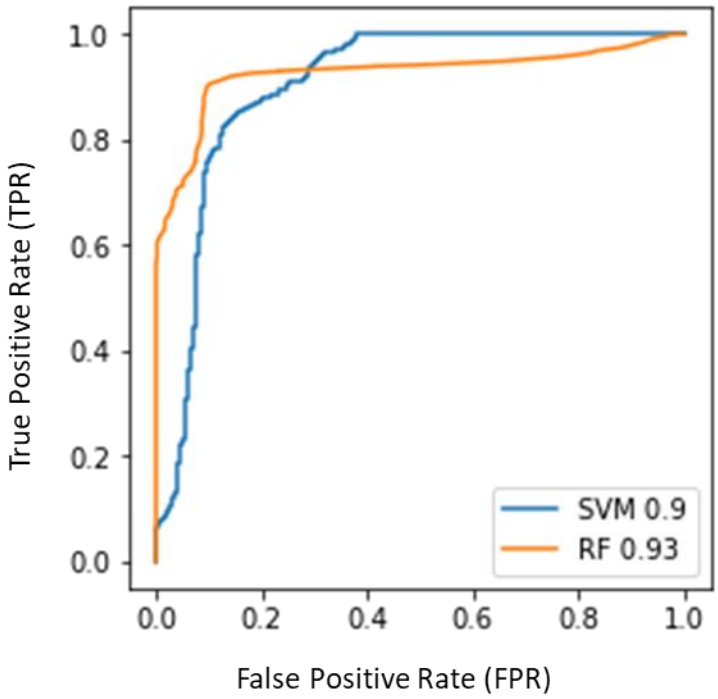


Figure 4.7 *Receiver Operating Characteristic (ROC) curve for the both the Support Vector Machine (SVM) and Random Forest (RF) classifiers, when detecting faults in video 1.*

Continuing to examine the confusion matrices for video 2, both classifiers perform considerably worse. Whilst FNRs have risen slightly to remain stable around 50%, FPRs have risen dramatically to 30% for the SVM and 38% for the RF. Again, this is likely due to the subtler 'Joint' faults prevalent throughout the sequence. Most interestingly, the ROC curves are quite unusual, intersecting at multiple points (Figure 4.8). Although neither curve entirely dominates the other, the SVM has a higher AUC of 74% compared to the 69% of the RF. This is the first instance of the SVM outperforming the RF, even though the raw frame by frame accuracy states otherwise. The discrepancy between these two statistics is due to the choice of threshold used to calculate accuracy.

Table 4.8 *Confusion rate matrix for the fault detection methodology, utilising a Support Vector Machine (SVM) classifier, applied to video 2.*

		Actual	
		Fault	Normal
Prediction	Fault	0.53	0.30
	Normal	0.47	0.70

Table 4.9 *Confusion rate matrix for the fault detection methodology, utilising a Random Forest (RF) classifier, applied to video 2.*

		Actual	
		Fault	Normal
Prediction	Fault	0.50	0.38
	Normal	0.50	0.62

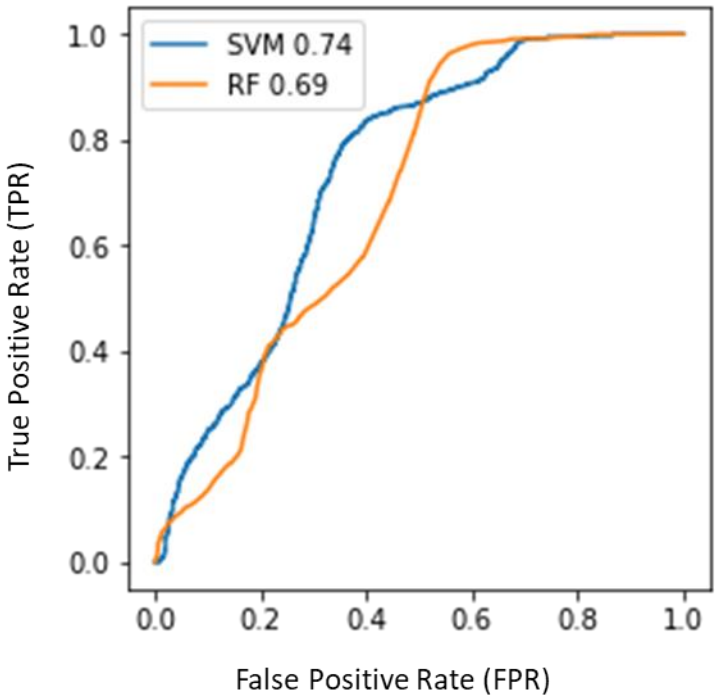


Figure 4.8 *Receiver Operating Characteristic (ROC) curve for the both the Support Vector Machine (SVM) and Random Forest (RF) classifiers, when detecting faults in video 2.*

Examining the final set of confusion matrices for the much longer video 3, both classifiers perform well, correlating with earlier results. The SVM achieved a FPR of 19% and a FNR of 21%, whilst the RF achieved a lower FPR of 13% and FNR of 17%. These are some of the best results for all three videos, balancing both

FPR and FNR. Continuing to analyse both ROC curves (Figure 4.9), the RF clearly dominates the SVM's curve, similar to those seen in Figure 4.5. Again, this domination is demonstrated by the RF's AUC of 84%, compared to the SVM's lower 68%.

Table 4.10 *Confusion rate matrix for fault detection, utilising a Support Vector Machine (SVM) classifier, applied to video 3.*

		Actual	
		Fault	Normal
Prediction	Fault	0.79	0.19
	Normal	0.21	0.81

Table 4.11 *Confusion rate matrix for fault detection, utilising a Random Forest (RF) classifier, applied to video 3.*

		Actual	
		Fault	Normal
Prediction	Fault	0.83	0.13
	Normal	0.17	0.87

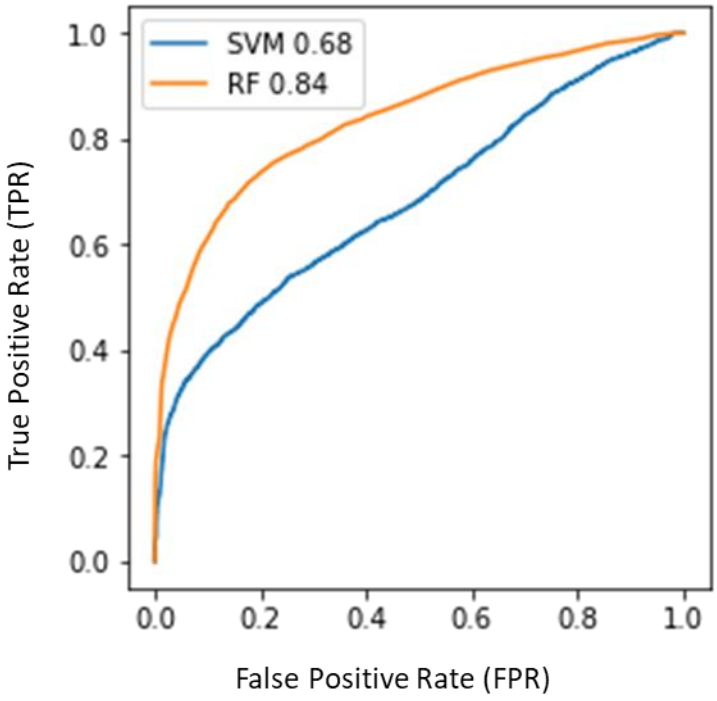


Figure 4.9 *Receiver Operating Characteristic (ROC) curve for the both the Support Vector Machine (SVM) and Random Forest (RF) classifiers, when detecting faults in video 3.*

Even though the fault detection methodology appears to perform well, especially when using the RF classifier, there are additional practical problems with the predictions. Given the motion of the camera and uncontrolled environment of most sewer surveys, both the SVM and RF's predictions tend to flicker, showing inconstancy within frames pertaining to a single fault. Whether through indecision (when a prediction lies on or close to the decision threshold) or occlusion of faults, this is not useful when labelling CCTV surveys. This issue is clearly demonstrated by Figure 4.10, which shows the predictions of each technique (red) compared with the surveyor's labels (blue) over the duration of Wessex Water video 1. Unlike the surveyor, both techniques (especially the SVM) are unable to clearly define blocks of frames containing a fault. The intermittent classification could be caused by a combination of multiple factors:

- The decision threshold of 0.5 could be too high, creating indecision where a fault prediction lies on or very close to the threshold. The intermittent classification appears to be most prevalent towards the start/end of a fault, where a decision threshold would be first crossed by a fault. This theory is supported as both classifiers show no indecision when classifying the fault between frame 2300 and 3300.
- The surveyor's annotations are at too low a resolution, resulting in labels that contain multiple smaller faults. Given the block like nature of all the surveyors' annotations, this theory is also viable, although some faults are labelled on a low resolution of around 100 frames, which correlates to roughly 0.25m of pipe. This point is further reinforced by the lack of surveyor labels for some smaller faults as shown in Figure 4.6.
- Faults could be present, although not visible due to the motion of the camera/occlusion. All three surveys were recorded with pan/tilt/zoom functionality. As such, a surveyor may annotate a fault whilst the camera is moving, and the fault is not currently visible. This appears to be most common whilst a surveyor is manoeuvring the pig to perform additional analysis of a fault.

It is likely that all three of the above points are factors affecting the flickering nature of the predictions. Given two of the three points are not controlled by the

methodology, an additional smoothing stage has been developed to minimise this issue (see Chapter 5).

A final noteworthy point that differentiates the application of fault detection on still images and continuous video footage, is the distribution of faulty frames. Both systems are calibrated using the same Wessex Water dataset of still images which has a roughly 60:40 split of normal images to faults. However, the video sequences have a variety of normal to faulty frame distributions, often dominated by images of normal pipe. Although this does not appear to have affected the performance of the fault detection technique, it may negatively impact accuracy in videos with a larger bias towards normal or faulty frames.

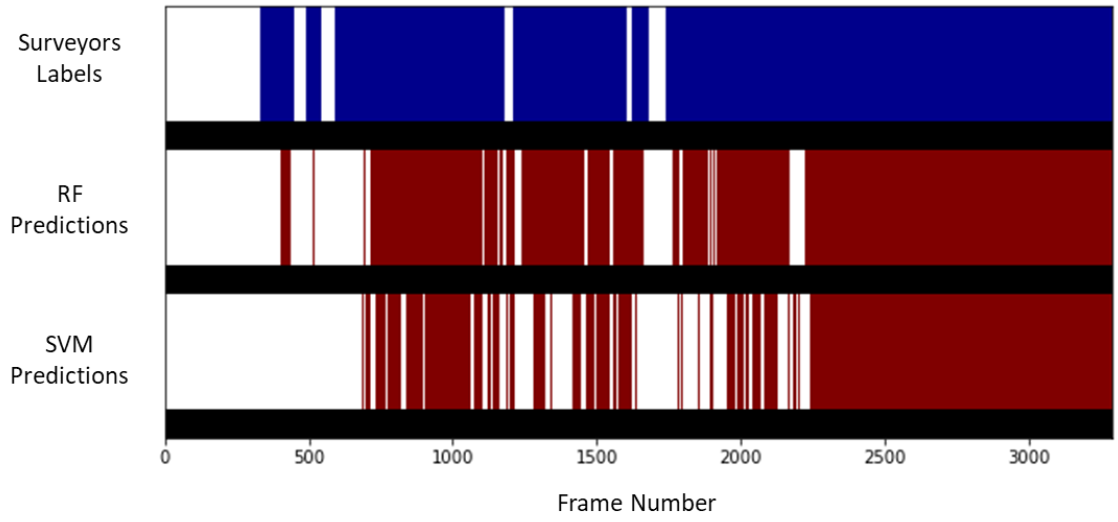


Figure 4.10 *A sequence plot of the Random Forest (RF) and Support Vector Machine (SVM) predictions against the surveyor's annotations, in Wessex Water video 1. White space indicates normal frames, whilst blue indicates a surveyor's label and red indicates a predicted fault.*

To summarise, the fault detection methodology performs well on continuous CCTV footage. The RF usually outperforms the SVM, repeatedly achieving a higher accuracy and better performance in terms of AUC. However, as highlighted in the first case study (section 4.3) subtler faults are not always detected. This is clearly demonstrated by the worse performance of both techniques in Wessex Water video 2, which contained many subtle 'Joint' defects. Furthermore, due to multiple factors, predictions can be inconsistent across the duration of a fault. This can make a single fault appear as multiple smaller faults, slowing annotation and causing unnecessary confusion about the volume of

defects within a pipe. To counter these inconsistencies a new smoothing stage is developed and applied in Chapter 5.

4.5 Summary

In conclusion, this chapter has given details of the fundamentals of the fault detection methodology and evaluated its effectiveness on still images and continuous CCTV footage, showing that it generally performs well although it fails to detect some of the subtler faults. The four stages of the method have been described in detail, and default parameters have been provided for the methodology's future developments.

The first case study demonstrates the methodology when applied to individual still images taken from CCTV surveys. The tool was applied to the core Wessex Water dataset (section 3.2.1) using both a Random Forest (RF) and Support Vector Machine (SVM) classifiers. These techniques achieved accuracies of 83.0% and 73.7% respectively, proving their usefulness for the fault detection task.

The second case study demonstrates the methodology, using both classifiers, on a collection of three unseen video sequences. Although the detection methodology generally performed well, with a best accuracy of 84.7% using a RF, it requires an additional smoothing stage to eliminate the presence of flickering in its predictions.

The next chapter (Chapter 5: Extended Fault Detection), examines extensions and improvements to the tool. These improvements include the addition of a smoothing stage and the intuitive combination of classifiers' predictions to improve accuracy.

Chapter 5: Extended Fault Detection Method

5.1 Introduction

Chapter 4 demonstrated the potential of the original fault detection method. Utilising GIST feature descriptors and a Random Forest (RF) classifier, the tool achieved a peak accuracy of 84.7% when applied to continuous CCTV footage. However, the methodology, although successful, left room for further refinement.

Predictions made by both the RF and Support Vector Machine (SVM) classifiers, tended to be intermittent, quickly alternating between 'fault' and 'normal' predictions, often at a resolution of 1-5 frames. This trait would be impractical in applications, raising questions about the true location of a fault. Eliminating this flicker would not only improve the methodology's usefulness to engineering practice but would also likely improve the overall accuracy of the technique. To achieve this, a smoothing stage (section 5.2) has been developed, which can be appended to the current fault detection methodology to combat this flicker.

After further investigation into the misclassifications of both the SVM and RF classifiers, it is clear that both perform well in distinct areas. Therefore, if the predictions of each could be intuitively combined, the methodology's accuracy could be improved, combining the best of each classifier. This is possible as both classifiers produce misclassifications unique to each technique. To achieve this a technique called stacking (section 5.4) was implemented in the extended method, utilising an additional machine learning classifier to combine the predictions of the RF and SVM.

Given the structure of the original fault detection methodology, a handful of additional tweaks and alternatives have been explored (section 5.6). These include: the implementation of an anomaly detection technique, utilising the structure of a RF to locate a fault within an image, and the implementation of a convolutional neural network (CNN).

Finally, in section 5.7, the Extended Fault detection methodology has been detailed. This finalised version of the tool details the recommended structure and parameters required for both on and offline implementation and use in the field.

5.2 Smoothing Methodology

Given the intermittent nature of classifications (section 4.4.2) performed by the original fault detection methodology, a new smoothing stage has been developed. This stage appends to the detection process and utilises two techniques: a Hidden Markov Model (HMM) and order oblivious filtering, also known as windowing. The aim of these technologies is not just to eliminate intermittent classifications but to improve detection rates by inferring additional information from the rest of the CCTV sequence.

The final stage of the methodology smooths the entire sequence of predictions by updating each frame's prediction based on those of its neighbours. The overall aim of this stage is to reduce the impact of noise and the camera's motion on detection accuracy, eliminating isolated inconsistencies in predictions to improve the detection rate on continuous footage. The smoothing is accomplished by first using a Hidden Markov Model (HMM) (Rabiner, 1989) and then Order Oblivious Filtering (Yan, Chakraborty, Misra, Jeung, & Abere, 2012). The insertion of the smoothing stage into the original fault detection method is shown in Figure 5.1. While unnecessary in previous experiments, which used a dataset of isolated still images, the HMM incorporates information from neighbouring frames to improve prediction rates.

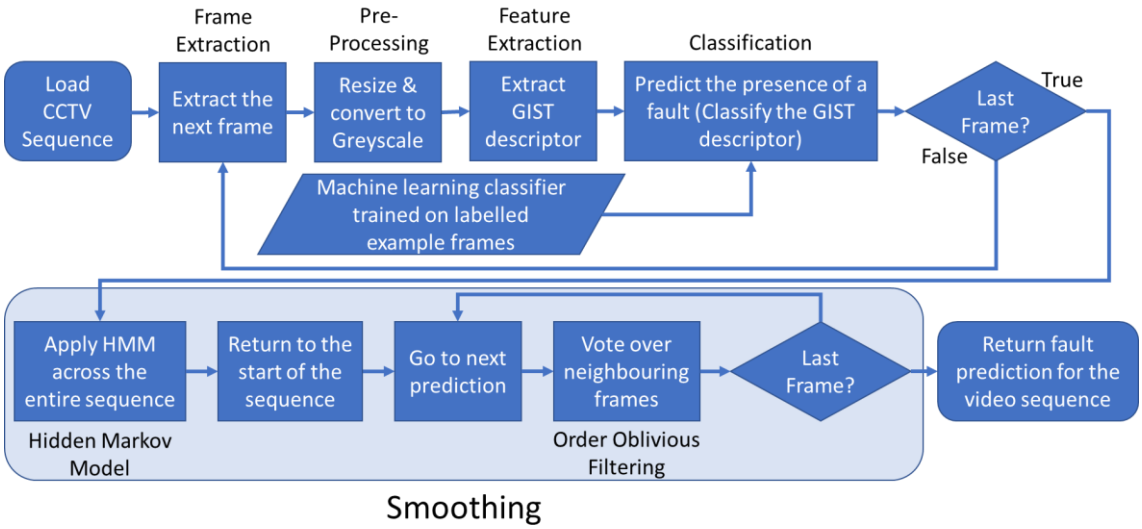


Figure 5.1 *Flowchart outlining the inclusion of smoothing in the fault detection method.*

In a Hidden Markov Model framework each frame's faulty/normal state is regarded as a hidden variable; what is instead observed is the video frame itself. The probability of observing the GIST features in each of the faulty and normal states is modelled here by the trained classifier (SVM or RF). Employing a Markov model makes the common simplifying assumption that each hidden state depends only on the state of the previous frame, and not on all the preceding frames. In the absence of observations, the probability of the hidden state remaining the same as the preceding frame or changing is governed by the state transition matrix.

This matrix can be manually estimated by examining the state changes in similar CCTV footage, or automatically calculated using the Baum-Welch algorithm (Rabiner, 1989). In this methodology, the Baum-Welch algorithm uses the expectation-maximisation algorithm (Dempster, Laird, & Rubin, 1977) to find the maximum likelihood parameters for the state transition matrix, given the estimated probability of each frame's state from the observations. Once the model parameters (the state transition matrix) have been learned, the HMM may be used to infer the hidden state from the observed GIST features. This may be achieved either in an online fashion, using only the current frame and the preceding frames (known as *filtering*), or using the forwards-backwards algorithm which incorporates information from frames following the current frames as well as preceding frames, resulting in a smoother prediction of the sequence's states and fewer anomalous transitions (Rabiner, 1989). In the case study (section 5.4), smoothing was performed retrospectively, using the forwards-backwards algorithm over the entire sequence to be classified. However, to permit close to real-time performance filtering may be used, or since only a relatively few frames ahead have a significant bearing on the current frame's state, smoothing can be done shortly after each frame has been recorded.

After smoothing with the HMM, the classifications were further filtered using a sliding window, also known as Order Oblivious Filtering (Yan, Chakraborty, Misra, Jeung, & Abere, 2012) or Windowing. This simple process aims to eliminate any further isolated inconsistencies in the predictions. It is accomplished by assigning a frame's predicted state to be the state of the majority of frames in a symmetrical window of W frames preceding and following it. By performing this last step, a

predicted state (‘faulty’ or ‘normal’) is assigned to every frame, utilising as much information as possible provided by the original video sequence.

5.3 Case Study: Smoothing

This case study investigates the inclusion of the new smoothing stage within the original fault detection method. Applied to three previously unseen CCTV videos, the process utilises a Random Forest (RF) classifier trained on the core Wessex Water dataset (section 3.2.1). The case study compares the performance of the original fault detection method, with the inclusion of the Hidden Markov Model (HMM) and then the inclusion of the entire smoothing stage.

5.3.1 Data and Setup

To demonstrate the new smoothing stage, the original fault detection methodology has been calibrated on the core Wessex Water dataset, utilising the default parameters given in Table 4.1. The RF classifier used by the methodology contains 100 trees and the same predictions have been utilised to evaluate the original methodology, HMM and full smoothing stage. The three unseen video surveys utilised in this case study are the same as those used in section 4.4; a summary of the videos can be seen in Table 4.5. As the smoothing stage has been included to interpret a frame’s predicted probability of containing a fault, only the original methodology utilises a threshold of 0.5 for identifying a frame as faulty. Even so, the internal parameters of the smoothing stage can be varied in order to generate Receiver Operating Characteristic (ROC) curves to evaluate all three experiment at all threshold values. For a full discussion about the significance of ROC curves see section 4.3.2.

5.3.2 Results and Discussion

To compare the improvements of each step in the smoothing stage, the frame by frame accuracy was initially calculated for all three surveys. Application of the smoothing achieved improved detection accuracy in video 2 from 53.0% (when applying the original methodology) to 59.3% after the application of the HMM and 59.9% once the smoothing stage including order oblivious filtering had been completed. Conversely, the frame by frame accuracy appeared to deteriorate after smoothing in videos 1 and 3. The original methodology achieved an accuracy of 84.7% in video 1, whilst the HMM reduced this to 69.3% and

completing the smoothing stage continued to reduce this to 69.2%. The same, although less extreme, occurred in video 3, with the original fault detection tool achieving an accuracy of 83.6% and the inclusion of the HMM and all smoothing reduced this to 82.3% and 82.5% respectively.

Continuing to examine the confusion rate matrices (Table 5.1 and 5.2) for video 1, the worst-case scenario, the effects of smoothing are clear for a threshold of 0.5. Whilst the true negative rate (TNR) remains stable around 1.0, the true positive rate (TPR) is reduced by the application of smoothing. The TNR is the rate at which normal pipe is correctly identified, whilst the TPR is the rate at which faults are correctly identified. This reduction in TPR or increase in false negative rate (FNR) correlates with the reduction in accuracy.

Table 5.1 *Confusion rate matrix for the original fault detection methodology, utilising a random forest (RF) classifier, applied to video 1.*

		Actual	
		Fault	Normal
Prediction	Fault	0.51	0.02
	Normal	0.49	0.98

Table 5.2 *Confusion rate matrix for the original fault detection methodology, utilising a random forest (RF) classifier and the new smoothing stage, applied to video 1.*

		Actual	
		Fault	Normal
Prediction	Fault	0.33	0.00
	Normal	0.67	1.00

However, the results presented in Table 5.1 and 5.2 only utilise a classification threshold of 0.5. The use of a single threshold does not demonstrate the true impact of the smoothing stage. This impact is better shown in Figure 5.2, in which the ROC of the smoothed methodology dominates that of the original tool at most thresholds. This is confirmed by the areas under the curve (AUC), where the original methodology and the inclusion of the HMM achieve an AUC of 0.93, whilst application of the smoothing stage increases this to 0.95.

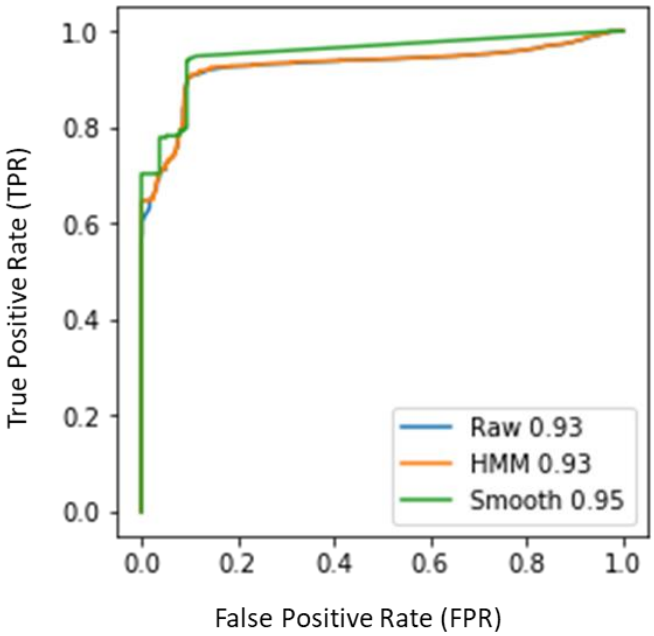


Figure 5.2 *Receiver Operating Characteristic (ROC) curve for the original fault detection methodology, the inclusion of the HMM, and full smoothing stage, when detecting faults in video 1.*

The relevance of threshold choice is further reinforced by example video 2. As mentioned in section 4.4.2, this video contains many of the subtler 'Joint' faults incurring the most misclassifications by the methodology. Comparing the confusion matrices of the original method (Table 5.3) against that of the smoothed (Table 5.4), show an improved TPR and TNR. However, examining the ROC curves (figure 5.3), shows that the smoothed methodology only dominates the original methodology (and HMM application) for a minority of thresholds and achieves an AUC 0.04 lower than the original methodology's AUC of 0.69.

Table 5.3 *Confusion rate matrix for the original fault detection methodology, utilising a random forest (RF) classifier, applied to video 2.*

		Actual	
		Fault	Normal
Prediction	Fault	0.50	0.38
	Normal	0.50	0.62

Table 5.4 *Confusion rate matrix for the original fault detection methodology, utilising a random forest (RF) classifier and the new smoothing stage, applied to video 2.*

		Actual	
		Fault	Normal
Prediction	Fault	0.56	0.33
	Normal	0.44	0.67

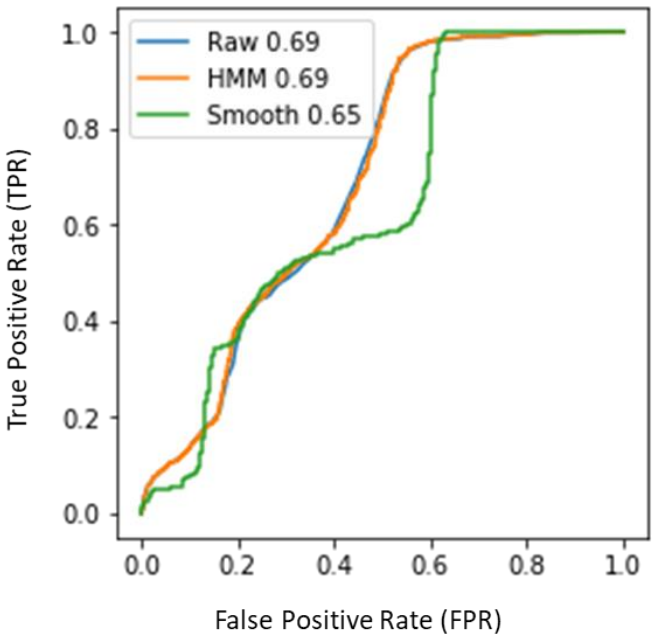


Figure 5.3 *Receiver Operating Characteristic (ROC) curve for the original fault detection methodology, the inclusion of the HMM, and full smoothing stage, when detecting faults in video 2.*

Analysing the third and final video sequence, the smoothing methodology performs differently again. The confusion rate matrices for the original (Table 5.5) and smoothed (Table 5.6) methodology show the smoothing reduces the TNR, incurring more false positives (FPR). Although this is better than an increase in FNR, implying that faults are missed, the additional (if minimal) time required to analyse normal pipe is not ideal.

Table 5.5 *Confusion rate matrix for the original fault detection methodology, utilising a random forest (RF) classifier, applied to video 3.*

		Actual	
		Fault	Normal
Prediction	Fault	0.83	0.13
	Normal	0.17	0.87

Table 5.6 *Confusion rate matrix for the original fault detection methodology, utilising a random forest (RF) classifier and the new smoothing stage, applied to video 3.*

		Actual	
		Fault	Normal
Prediction	Fault	0.90	0.38
	Normal	0.10	0.62

Yet again the relevance of the threshold choice is demonstrated by the experiment's ROC curves (Figure 5.4). As with video 1, the smoothed methodology dominates the original methodology for most threshold values, achieving a higher TPR for a lower FPR. This is reinforced by the increased AUC of 88% for the smoothed methodology over the original tools AUC of 84%. As such it can be said that, smoothing improves the detection accuracy, even if this contradicts with the perceived accuracy at a decision threshold of 0.5.

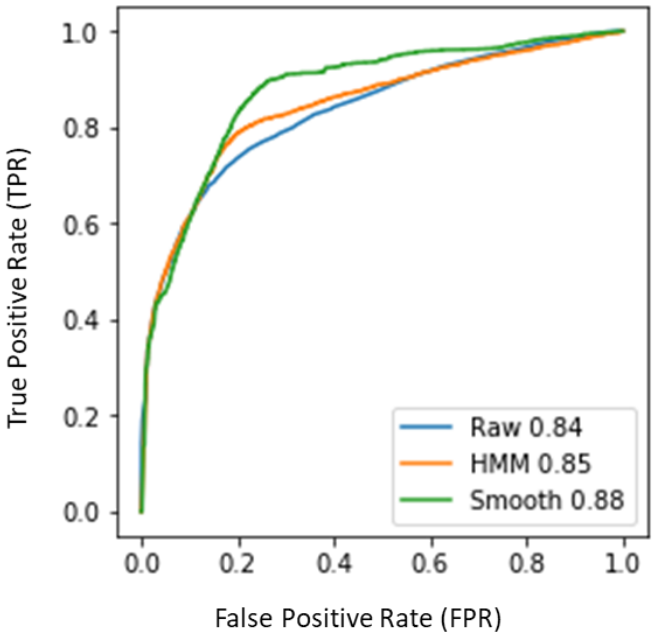


Figure 5.4 *Receiver Operating Characteristic (ROC) curve for the original fault detection methodology, the inclusion of the HMM, and full smoothing stage, when detecting faults in video 3.*

Having thoroughly investigated the impact of smoothing on the accuracy of the fault detection method, the impact on flickering was also explored. As was intended, the implementation of smoothing did reduce the intermittent classification of the system as can be seen in the sequence plot of video 2 (Figure 5.5). Although it does not directly address the suspected causes, such as annotation resolution or occlusion (see section 4.4.2), smoothing has provided a much more practical set of fault labels. As such the method would be of much more use to an engineer analysing the survey.

In addition, it is worth noting that although the accuracy was low for video 2, each of the surveyor's annotations has been identified by the tool. The loss in accuracy actually appears to be associated with the start/end of a fault, which as noted by (van der Steen, Dirksen, & Clemens, 2014) is a surprisingly subjective topic.

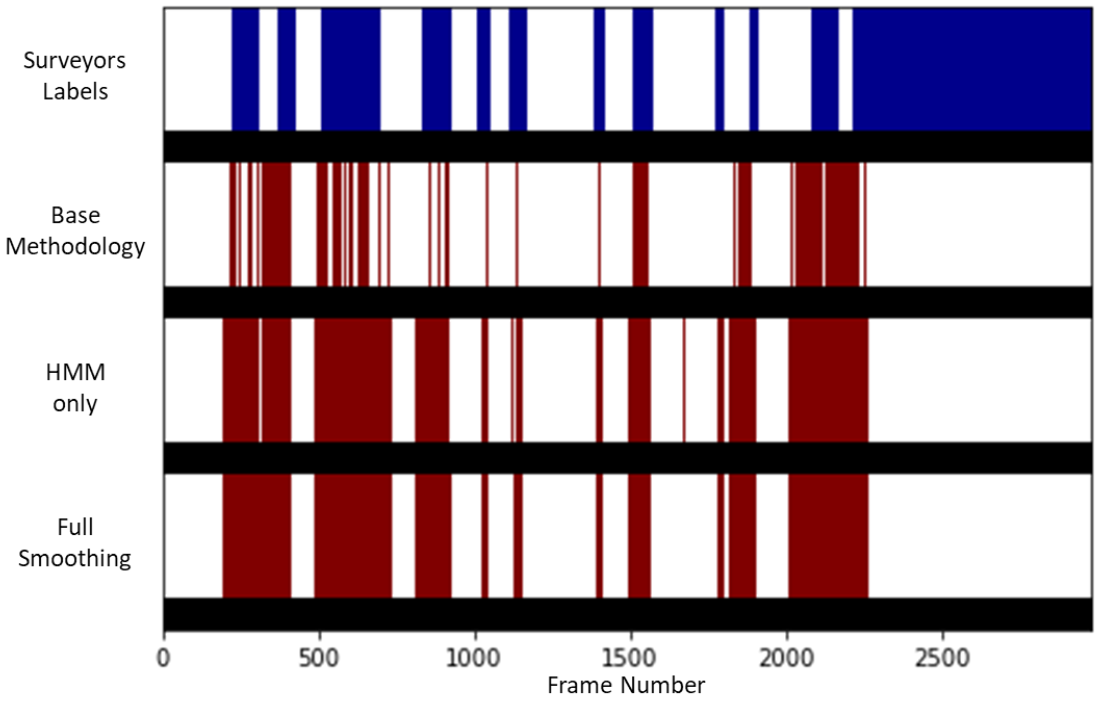


Figure 5.5 *Sequence plot of the original methodology's, HMM's and smoothing stage's predictions against the surveyor's annotations in video 1. White space indicates normal frames, whilst blue indicates a surveyor's label and red indicates a predicted fault.*

5.3.3 Summary

In summary, the smoothing stage is a valuable addition to the methodology, often demonstrating an improved accuracy and eliminating intermittent classifications. The only caveat to these improvements is that the choice of classification threshold becomes increasingly important. A discussion on this point is included in section 5.7.3. It is also worth noting that the inclusion of the smoothing stage had a negligible impact on the methodology's computational efficiency, meaning that it is still viable for real-time implementation.

5.4 Stacking

As demonstrated in the case studies (section 4.3), both the Random Forest (RF) and Support Vector Machine (SVM) are capable classifiers, performing well as part of the overall fault detection pipeline. However, by examining the shortcomings of each technique, it can be seen that each has a number of misclassifications unique to the particular classifier. From Table 5.7 it can be seen that up to 18.8% of misclassifications could have been avoided if the alternative classifier was used. Therefore, the stacking methodology presented here aims to

combine the predictions of multiple machine learning classifiers to improve the accuracy of the fault detection (see Figure 5.6 for a schematic outline). By doing so, each classifier should compensate for the other's shortcomings, i.e. the SVM's improved detection of normal pipe will complement the RF's detection of faults.

Table 5.7 *Table displaying the number of unique misclassifications for the Random Forest (RF) and Support Vector Machine (SVM) classifiers when applied in the fault detection methodology to the Wessex Water dataset of still images. The misclassifications have been broken down by frame contents. '# Unique' refers to the number of frames misclassified solely by the respective classifier. '%' refers to the percentage of that that categories frames uniquely misclassified.*

Fault	SVM		RF	
	# Unique	%	# Unique	%
None	113	2.0	202	3.7
Multi	23	7.5	8	2.6
Deposits	24	6.3	15	3.9
Surface	8	3.3	4	1.7
Joints	59	7.7	15	2.0
Obstructions	6	8.8	3	4.4
Crack	20	8.3	7	2.9
Roots	16	8.0	0	0.0
Infiltration	11	9.5	8	6.9
Hole	1	7.5	0	0.0
Deformation	0	0.0	0	0.0
Broken	6	18.8	0	0.0
Brickwork	0	0.0	0	0.0
Other	4	11.8	2	5.9

To intelligently combine each classifier's predicted probabilities, the stacking technique has been implemented (Sill, Takács, Mackey, & Lin, 2009). This technique implements a second level classifier, which combines the predictions of multiple models. In this work, a SVM has been selected as the second level classifier, due to its computational efficiency and previous successes. Taking only the predicted probabilities of the primary classifiers as inputs, the stacking classifier produces a final predicted probability considering those of the independent classifiers.

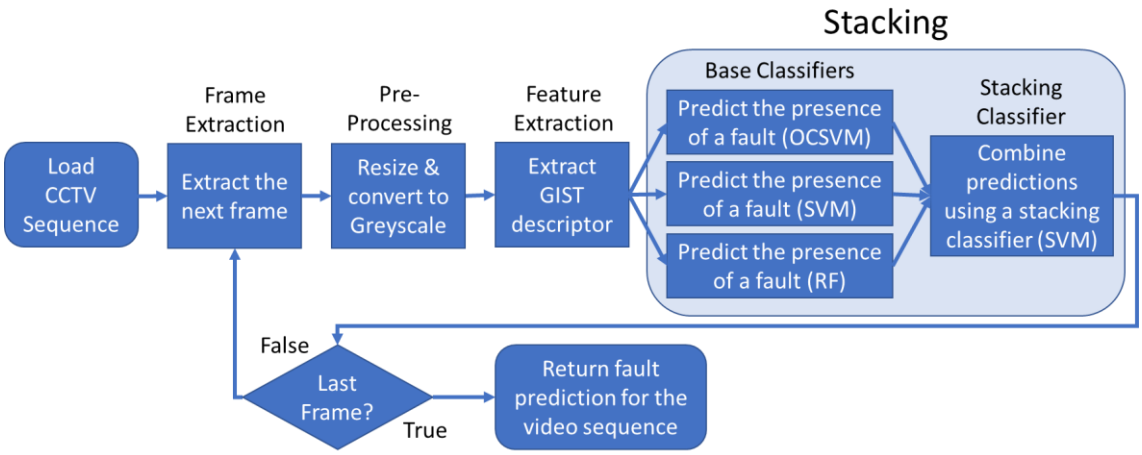


Figure 5.6 *Workflow for the original fault detection methodology with the inclusion of the stacking technique.*

An extension of this technique, called feature-weighted stacking (Sill, Takács, Mackey, & Lin, 2009), has also been applied and examined in a case study (section 5.5). In addition to the predictions of the base classifiers, the secondary stacking classifier also utilizes the GIST feature descriptor. Doing so provides the stacking classifier with all possible information about the frame, with the intention of improving the final prediction. However, this extra information makes the final stacking classification problem more complex, diluting the predictions of the primary classifiers. This may result in minimal improvement, as the stacking classifier performs similarly to the base classifiers.

In both versions of this technique, the stacking classifier is trained on a collection of labelled images. Each image should have both a label, and a predicted probability from each primary classifier, which means these images need to be different from those used to train the base classifiers.

5.5 Case Study: Stacking

This case study investigates the integration of the stacking technique with the fault detection methodology (see Figure 5.6). Applied to the three unseen CCTV video sequences, stacking combines the predictions of a Random Forest (RF), a Support Vector Machine and a One-Class SVM (OCSVM) (Chen, Zhou, & Huang, 2001). See section 5.6.1 for additional insight into the OCSVM. The most prominent results of these experiments are identified and compared, to identify the benefits of the application of stacking and feature-weighted stacking.

5.5.1 Data and Setup

To test the effect of stacking on the methodology, a RF, SVM and OCSVM classifier have been calibrated using the core Wessex Water dataset, utilising the parameters given in Table 4.1. However, unlike previous case studies, an additional RF, SVM and OCSVM were trained using only two thirds of the dataset, calibrating both the stacking and feature-weighted stacking classifiers on the remaining third. This ensured that no image was used to calibrate both the base classifiers and stacking classifier. After preliminary experimentation, SVMs were chosen as the stacking classifiers as they achieved the highest accuracies. These were simple SVMs, using a linear kernel and a regularisation coefficient (C) of 10, which was selected using cross validation.

As before, the techniques are demonstrated on three example video sequences, a summary of which can be seen in Table 4.5. The techniques are compared using raw accuracy at a prediction threshold of 0.5 and the respective Receiver Operator Characteristic (ROC) curves.

5.5.2 Results and Discussion

Comparing frame by frame accuracies, stacking-based detection methodologies have mixed successes. Applied to the first video sequence, the stacking techniques achieve accuracies of 77.7% and 76.3% for stacking and feature weighted stacking respectively. These are higher than the accuracies of both the RF (76.9%) and SVM (55.0%), although lower than the One-Class SVM (OCSVM) (85.6%). A similar story can be told for both video 2 and 3, each achieving reasonable accuracies, if not the best (see table 5.8 for a breakdown).

Table 5.8 *Table presenting the accuracy and area under the curve (AUC) metrics for classifiers applied to three test video segments. The Random Forest, Support Vector Machine and One-Class Support Vector Machine have been presented alongside the stacked and feature-weighted stacked combinations. Accuracy and AUC have been given as a percentage (%), whilst bold figures indicate the best method for each video and performance measure.*

Technique	Video 1		Video 2		Video 3	
	Acc.	AUC	Acc.	AUC	Acc.	AUC
Random Forest (RF)	76.9	94	54.2	68	**83.4**	**81**
Support Vector Machine (SVM)	55.0	90	51.0	67	77.3	64
One-Class Support Vector Machine (OCSVM)	**85.6**	76	**71.0**	72	75.2	52
Stacking	77.7	**95**	51.7	**69**	83.2	80
Feature Weighted Stacking	76.3	**95**	51.8	**69**	82.0	79

In the same vein, the stacking techniques tend to achieve a higher AUC, than most of their constituent classifiers, if not all. This is most evident in video 1, where both stacking techniques achieve an AUC of 95%. This ‘good’ performance is further demonstrated by the technique’s ROC curves (see Figures 5.7, 5.8 & 5.9).

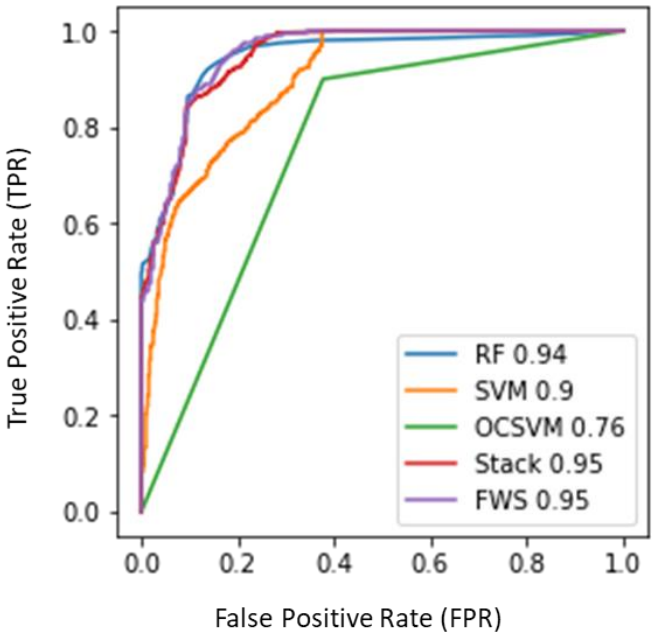


Figure 5.7 *Receiver Operating Characteristic (ROC) curve for video 1. Three fault detection methods based on independent classifiers (RF, SVM and OCSVM) are compared with the two fault detection methods based on stacking methodologies (Stacking and FWS).*

Applied to video 1, both stacking approaches' ROC curves contest that of the RF classifier. As it can be seen from Figure 5.7, across the full range of prediction thresholds, stacking performs slightly better than any of the independent classifiers, effectively selecting the best classifier for each frame.

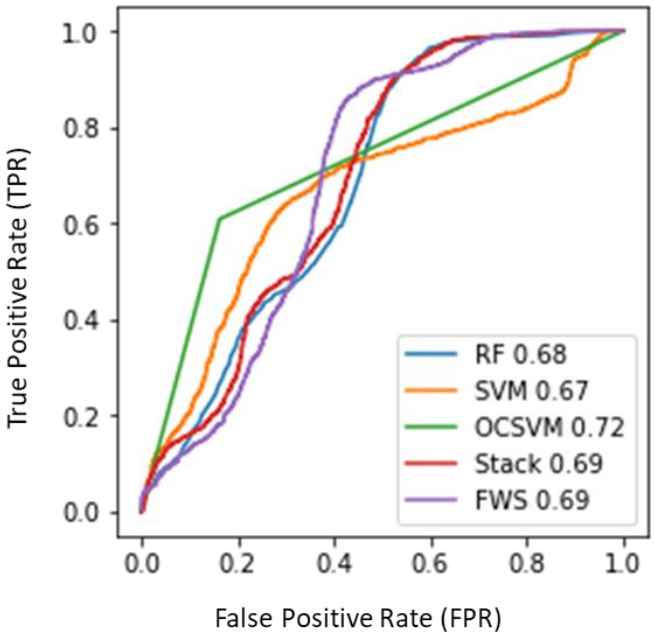


Figure 5.8 *Receiver Operating Characteristic (ROC) curve for video 2. Three fault detection methods based on independent classifiers (RF, SVM and OCSVM) are compared with the two fault detection methods based on stacking methodologies (Stacking and FWS).*

As can be seen from Figure 5.8, the performance of stacking is much harder to define in video 2, achieving a slightly lower AUC than the OCSVM. Even so, no other techniques clearly dominate either of the two stacking ROC curves. In fact, feature weighted stacking clearly dominates all other techniques between a false positive rate of 0.3 and 0.4. This improvement is not prevalent in either the AUC or accuracy and demonstrates the importance of examining each ROC plot.

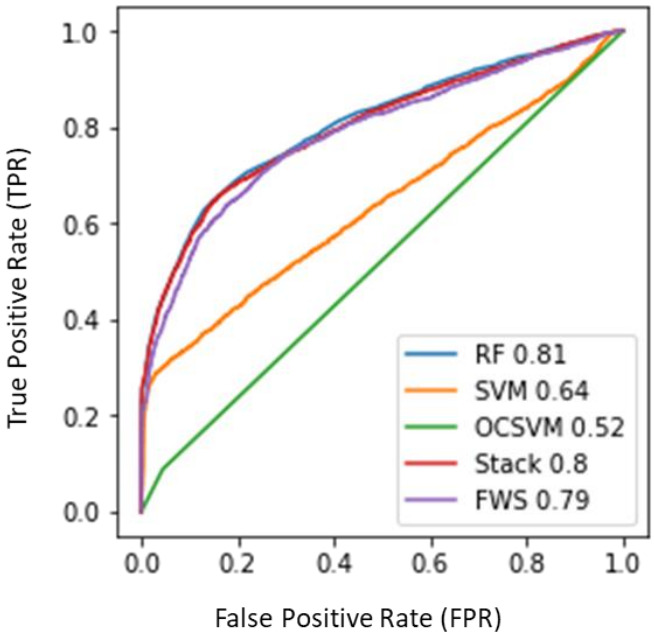


Figure 5.9 *Receiver Operating Characteristic (ROC) curve for video 3. Three fault detection methods based on independent classifiers (RF, SVM and OCSVM) are compared with the two fault detection methods based on stacking methodologies (Stacking and FWS).*

Finally, analysing the ROC curves achieved in video 3 (Figure 5.9), show the stacking classifiers again performing similarly to the RF. Although hard to distinguish, the RF achieves a marginally higher AUC of 81% to the stacked classifier's 80% and the feature weighted stacked classifier's 79%.

Even though the stacked classifiers perform well, with feature weighted stacking usually performing the best, they do not always appear to work as intended. The reason for this is that instead of selecting the most appropriate classification for each frame, stacking often produces an averaging effect. This is clear when analysing the ROC curves, where the stacked methodologies regularly appear to lie between the best and worst classifiers. This is not necessarily a bad thing as the stacking classifiers appear to perform in a more robust manner than the SVM and OCSVM. This stability is extremely desirable when working in the uncontrolled environment of sewer surveys. However, with a few exceptions, the RF performs almost as well as either stacked classifier, for considerably less computational effort. As such the application of stacking may be useful in only the offline implementation of the fault detection methodology.

5.5.3 Summary

To summarise, both stacking and feature-weighted stacking can be valuable additions to the fault detection methodology. In this case study stacking did not provide marked performance improvements, but improved the methodology's robustness, averaging between the predictions of three independent classifiers. Even so, some small improvements were achieved, often boosting the AUC of any given classification technique. Given stacking did not perform quite as intended, further investigation will include a larger variety of base classifiers.

5.6 Additional Experiments

Throughout the development multiple preliminary experiments have also been completed, each investigating an alternate approach. First an anomaly detection approach was taken to distinguish normal pipe from all other surveying observations. This is demonstrated using a One-Class Support Vector Machine (OCSVM), and assumes that anything that is not normal pipe would be labelled by a surveyor (including faults, and pipe features). The second experiment investigates the ability to trace decisions through a Random Forest (RF) classifier. Following through these decisions and the dense structure of the GIST descriptor the most probable location of a fault within a frame can be identified. The third and final experiment investigates the use of a deep Convolutional Neural Network (CNN) as an alternative classifier to detect faults. Due to the data large data requirements of most deep learning, this approach to detection re-trains the final layer of a CNN pre-trained on an independent set of non-sewer images.

5.6.1 One-Class Support Vector Machine (OCSVM)

Previous implementations of the fault detection methodology shown here have applied the Random Forest (RF) and traditional Support Vector Machine (SVM) classifiers, both of which distinguish between two (or more) defined classes. As an alternative, the One-Class Support Vector Machine (OCSVM) has been tested (Chen, Zhou, & Huang, 2001). This machine learning technique is designed for anomaly detection, identifying only a single class. In the fault detection methodology, this has been tested in place of a standard classifier. The OCSVM is calibrated using a dataset of frames displaying only 'normal' pipe, identifying whether subsequent images contain 'normal' pipe or not. This is in direct contrast

to the RF and SVM methodologies which were trained on both images containing ‘normal’ and ‘faulty’ pipe to identify whether a frame belonged to the ‘normal’ or ‘faulty’ class.

The interior structure of the OCSVM is very similar to that of a traditional SVM, mapping the data to a higher dimensional plane using a (Radial Basis Function) kernel. However instead of attempting to split two classes of data using a linear hyperplane, the single class is surrounded by a hypersphere. When applied to unseen data the OCSVM maps it to the higher dimensional space, performing the classification based on the derived hypersphere. If the data lies within the hypersphere it is assumed to be similar enough to the ‘normal’ class that it is classified as ‘normal’, however if the data falls outside the hypersphere it is flagged as an anomaly, not ‘normal’ pipe, and assumed to contain a fault. As this form of classification detects all anomalies within the CCTV footage, not specifically faults, other regions of interest could be flagged, including pipe joints and manholes. However, current surveying practices require technicians to annotate footage with these additional details, so this is not considered an issue when identifying faults.

In experiments, the OCSVM has performed well, achieving results similar to and sometimes superior to the RF and SVM classifiers. When applied to the videos used in previous case studies (section 4.3, 4.4, 5.3 & 5.5) the OCSVM outperformed both classifiers in video 1 and 2, achieving a peak accuracy of 85.6% (before smoothing). However, the technique can be inconsistent, struggling in video 3, where it achieved the lowest (if respectable) accuracy of 75.2% and an AUC of only 0.52 which is almost equivalent to random allocation. Given this nature, it makes an ideal candidate for the stacking technique, able to complement both the RF and SVM classifiers (see section 5.5).

5.6.2 Fault location

The impressive performance of the Random Forest (RF) was achieved despite its simple structure which is one of the simplest of many modern machine learning classifiers. This already makes it attractive for use for application in the detection methodology. However, preliminary experiments suggest that the RF’s structure also enables the location of a fault within an image. This would not only help a

surveyor label any faults within CCTV footage but would require minimal additional computational processing.

As discussed in section 4.2.4, RFs are comprised of a number of simpler decision trees, relying on its ensemble nature to produce accurate predictions (Breiman, 2001). Each individual tree is simple to parse and can quickly be traced by hand, giving an insight into the decisions made by the classifier. As such the importance of each feature can be calculated based on its position within a tree and potential for information gain. By identifying a forest's feature importance's and tracing the decision path through each tree the most used features can be identified for any classification. Following this line of thought, a subset of the 512 GIST features can be identified as the most important in any classification. Given the structure of GIST descriptors these individual features can be traced back to regions of the original image, highlighting the grid cell likely to contain a fault.

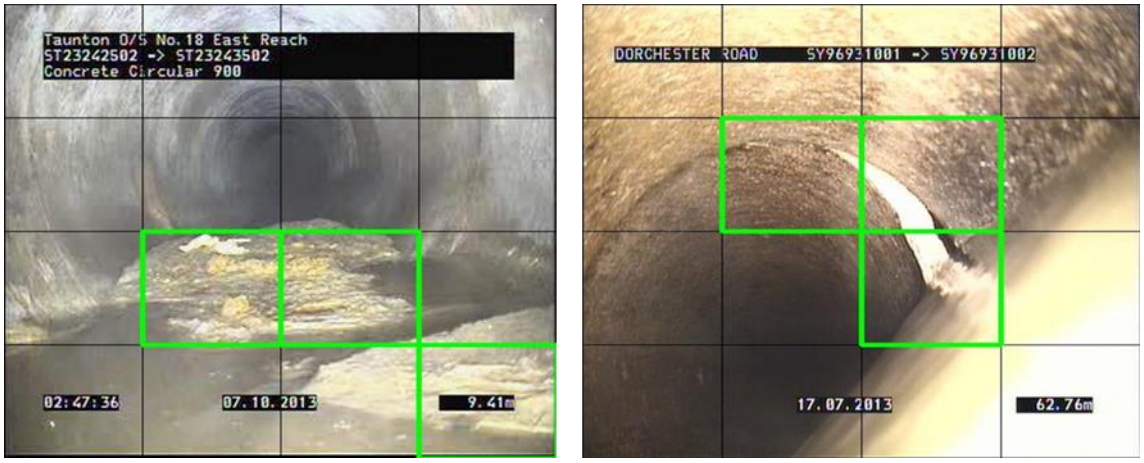


Figure 5.10 *Two example frames in which the location of the fault has been automatically identified using the feature importance extracted from the RF classification and structure of the GIST feature descriptor. Highlighted cells in the image show the locations of the most important features used in the detection of the faults in these frames.*

When applied to a subset of the core Wessex Water dataset, a number of faults were correctly identified, as shown in Figure 5.10. Although this has only been demonstrated in small scale preliminary experiments, this could be a distinct advantage of the RF classifier and would greatly benefit the fault detection methodology. Future investigation in larger scale experiments would evaluate the viability of this tool and help formulate an optimal process for identifying the most influential features in a fault's detection.

5.6.3 Convolutional Neural Network

As mentioned in section 4.3.2, a Convolutional Neural Network (CNN) was also applied to the fault detection problem. This machine learning tool is a class of deep learning neural network, commonly applied to problems in computer vision (Krizhevsky, Sutskever, & Hinton, 2012) (Simonyan & Zisserman, 2014) (Karpathy, et al., 2014). Inspired by the human understanding of eyesight, CNNs follow the structure of a traditional deep learning neural network (see section 2.6), applying a convolution operation to inputs at each neuron. Unlike a traditional neural network, inputs are weighted at each layer dependant on their relative locations. This means that even though every neuron is connected, local connections are the most influential (Krizhevsky, Sutskever, & Hinton, 2012). This complex structure enables a CNN to learn its own internal features from which to perform visual classifications. However, this does come at the cost of requiring a substantial dataset of images for effective calibration.

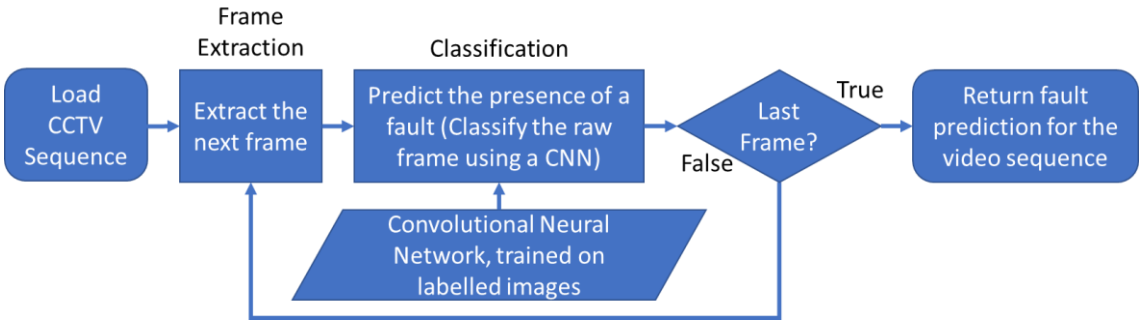


Figure 5.11 *Workflow for the application of a Convolutional Neural Network to the fault detection problem.*

In the context of fault detection, the CNN sits outside of the methodology defined in section 4.2.1 performing many of its steps internally (see figure 5.11). As such it has no need for GIST features or other classifiers. In simple experiments to evaluate the effectiveness of this tool, a pre-trained MobileNet (Howard, et al., 2017) CNN had its final layer re-trained and tested using the core Wessex Water dataset (see section 3.1.2) and 25-fold cross validation. This achieved an accuracy of 81%, comparable to that of the current fault detection methodology.

Although it achieved good results, further development with CNNs was eschewed for the approaches presented in this thesis for a few reasons:

- Deep learning neural networks generally require a large volume of data for best performance. At the time of development, data was limited, confining experiments to small scale demonstrations. As the volume of labelled data increases as does the viability of CNNs, as such CNNs are a clear option when more data is available.
- Neural networks are ‘black box’ classifiers. This means their internal decisions have very little perceptible logic, making it very hard for a human to reason through its decisions. Although not vital, the methodology is attempting to integrate with the well-established field of surveying. As such it is easier to convince practitioners to implement a transparent technology (Random Forest or Support Vector Machine), which they can reason through and hold to some level of accountability.
- Finally, the small-scale experiments did not demonstrate real-time application of the technology. This is not to say that real-time application of CNN’s is impossible but would require more thought than the current fault detection methodology.

Given these reasons, deep learning and CNNs are a clear area for further investigation, especially once additional labelled data is available. However, the small CNN that has currently been employed could be used as a base classifier when utilising the stacking methodology. Alternatively, a CNN’s internally calculated features could be tested in place of the currently applied GIST descriptor.

5.7 Extended Fault Detection Methodology

Having developed multiple additions to the fault detection methodology, this section aims to define its practical implementation. Using the original methodology detailed in section 4.2 as a base, each addition developed throughout the chapter is integrated and presented in a discussion of the tool’s practical online and offline applications. This discussion includes a description of the ideal setup, implementation and parameters for the methodology. It should be noted that these are not the only possible implementations. The methodology is flexible enough to add/alter stages with relative freedom, enabling an implementation tailored to the needs of every user.

5.7.1 General points

Regardless of the choice between online and offline implementation, many key requirements of the methodology remain the same. They are therefore discussed together here:

- Video footage to be analysed can be recorded using any device that creates standard CCTV footage. However, best results are achieved using a stable platform which does not utilise pan/tilt functionality, as this can often interfere with the perceived locations of faults. Due to the pre-recorded nature of the footage used in this thesis, the use of pan and tilt features was inevitable, however, future work would use dedicated surveys recorded without these features.
- The training process for all machine learning classifiers used throughout the methodology should be completed before application. This training should be performed on a labelled dataset of CCTV frames similar to those expected in future surveys. During analysis of the classification method (Chapter 6) it was found that at least 100 examples of each fault type were required for reasonable performance.
- The machine learning components can be re-trained on a regular (weekly or monthly) basis as the library of labelled frames grows. This assumes that the analysis of every survey can be appended to the dataset of images used for calibration. This is highly recommended, especially for surveys where the methodology is found to perform poorly.

5.7.1 Offline Implementation

Offline use of the fault detection tool implies the analysis of CCTV data after collection. It assumes the footage could be processed in an office or via a server before analysis by a technician. This processing could take part at any time of day, and whilst not requiring real-time application, it should not be slow to run as the analysis in this chapter demonstrates that the fault detection operates faster than 25 frames per second, the fastest rate at which frames are likely to arrive. Finally, for the best practices regarding the application of the methodology's results see Chapter 7.

Figure 5.12 displays the workflow of the offline methodology application, from initial feature extraction to the final predictions. As it can be seen from this figure, the original four stage structure remains, with the inclusion of a fifth stacking and final smoothing stage. As there is no requirement of real-time application, stacking is also performed, combining the predictions of a Support Vector Machine (SVM), Random Forest (RF) and One-Class Support Vector Machine (OCSVM). The additional machine learning classifiers (such as neural networks) could also be included in the stacking process. The additional fault location process (section 5.6.2) has been omitted from this proposal as it requires additional development and evaluation before implementation.

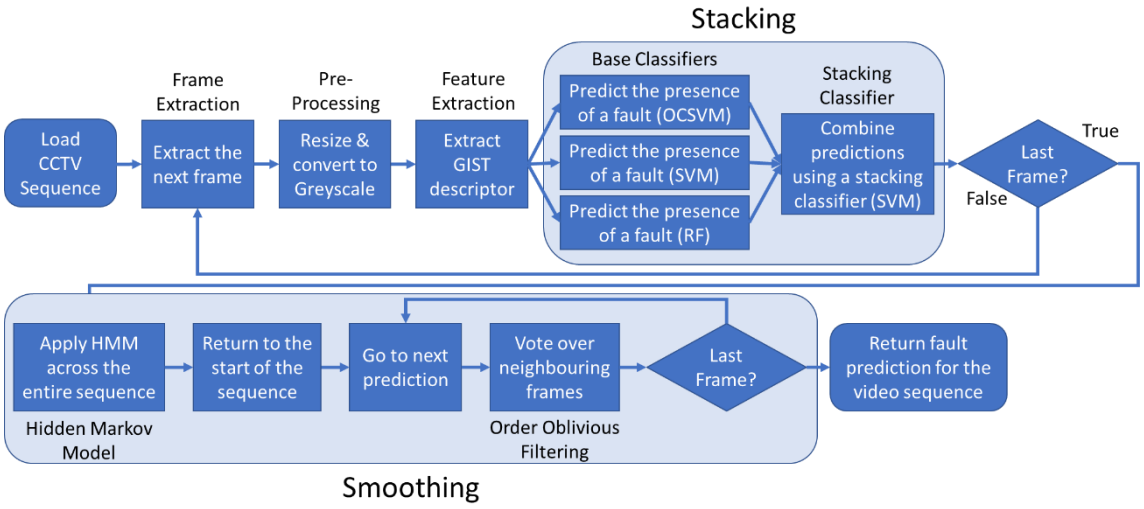


Figure 5.12 *Workflow for the offline implementation of the fault detection methodology.*

5.7.2 Online Implementation

Online implementation of the tool requires real-time calculations, capable of providing an overlay for a surveyor during the recording process. As such, speed and accuracy are paramount. The footage should be processed utilising the same computer used to annotate the survey, ideally a laptop or desktop computer that can be transported alongside the surveying equipment. This implementation could run as part of the recording process, processing images and identifying faults before passing the information to surveying software. Alternatively, it could run in parallel, processing images within the surveying software. Further discussion is given in chapter 7.

Figure 5.13 displays the workflow of the online methodology application. As with the offline implementation, the original four stage structure of the methodology remains the same (section 4.2.1), with the addition of the smoothing stage. However, stacking has been omitted, given it requires additional computational power. This omission of stacking is based solely on the un-optimised code currently being used by the method. Future online implementations would aim to include this stage, after the code has been re-written using more efficient programming paradigms and multi-threading. Smoothing is also performed dynamically, after the analysis of every frame, rather than after generating predictions for the entire survey. As the online implementation utilises only a single machine learning classifier, the Random Forest was selected for its reliable performance over all past case studies (sections 4.3, 4.4, 5.3 & 5.5). These factors may have a slight negative impact on performance, but the trade-off for real time application is deemed acceptable.

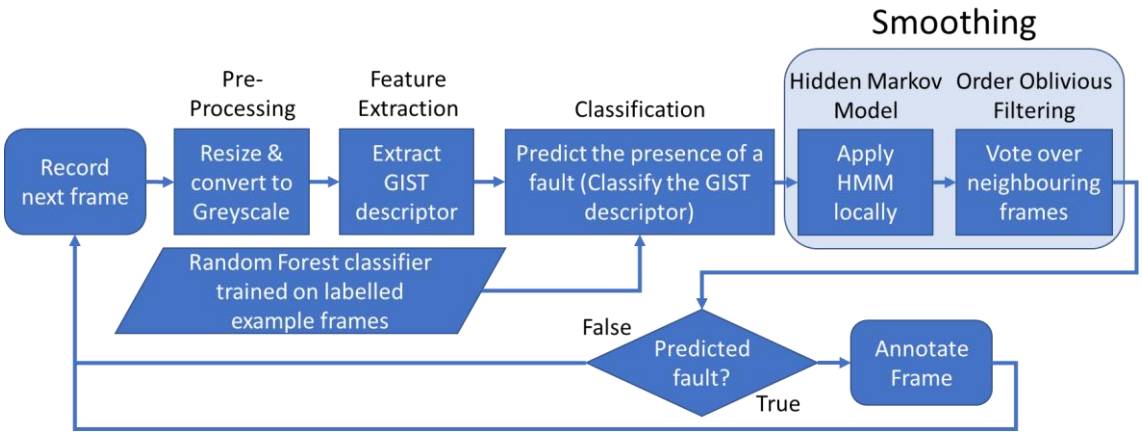


Figure 5.13 *Workflow for the online implementation of the fault detection methodology.*

5.7.3 Parameters

The key parameters of the methodology are provided in Table 5.9, with principal differences between the online and offline implementations identified. It is worth noting that the default parameters are only guidelines for future applications of the technology, and additional tuning via cross validation is always recommended.

Table 5.9 *A table defining the recommended (default) parameters for both online and offline implementations of the fault detection methodology.*

Parameter	Offline	Online	Description
F	25	25	The rate at which frames are extracted from the raw CCTV footage (fps). As most CCTV footage is recorded at 25 frames per second, it is recommended that all frames are processed, even during online application. The fps could be reduced if speed is of great importance, however this will negatively impact accuracy.
R	128x128	128x128	The resized resolution of extracted frames. This resolution is selected for both offline and online implementation as higher resolutions achieved very little increase in accuracy for much longer processing times.
Θ	8	8	Θ is the number of orientations of Gabor filter used to calculate a GIST descriptor.
S	4	4	S is the number of scales of Gabor filter used to calculate a GIST descriptor.
Cells	4x4	4x4	'Cells' is the resolution of the cell grid used to calculate a GIST descriptor. All three of these parameters are key to the calculation of the GIST descriptor and have been left as defined by Oliva & Torralba (Modelling the shape of a scene: A holistic representation of the spatial envelope, 2001). Preliminary experiments showed that increasing any of these parameters resulted in a small increase in accuracy for much longer processing times.
W	50	25	W is the number of frames considered when performing Order Oblivious Filtering (windowing). The default value for online application is half that of online as future frames would

Parameter	Offline	Online	Description
			not have been recorded. As such only previous neighbouring frames can be smoothed over.
Thresh.	0.5	0.5	'Thresh.' is the decision threshold selected to identify the presence of a fault. A default value of 0.5 has been shown to work well throughout the case studies presented in this thesis. However, this value should be selected according to the criticality of the surveyed pipe. The threshold can be set using the ROC curve to the desired true positive rate or the acceptable false positive rate. See section 7.2.3 for further details.
RF	No. Trees: 100		The Random Forest classifier can be applied for both online and offline fault detection. Its key parameter, the number of trees, is defined here. Increasing the number of trees beyond 100, provides little improvement in accuracy whilst increasing processing times.
SVM	C: 150 Kernel: RBF γ: 0.002		The Support Vector Machine is applied during offline fault detection but could be applied in place of the RF during online detection. The recommended regularisation constant (C), length scale (γ)and kernel type are given here, although polynomial kernels are also found to perform well.
OCSVM	ν: 0.001 Kernel: RBF γ: 10		The One-Class Support Vector Machine is used during offline detection but could also be applied in place of the RF during online detection. ν represents the contamination of the training data. γ is the kernel coefficient for the RBF kernel.
Stacking SVM	C: 10 Kernel: Linear		When used as a stacking classifier, the SVM should be much simpler, given the reduced number of parameters. As such, a smaller regularisation constant (C) of 10 is provided for a simple linear kernel.

5.9 Summary

In conclusion this chapter gives details of a number of additions to the fault detection methodology, improving its effectiveness for application in the real world. Smoothing was shown to be an effective tool, including the use of a Hidden Markov Model and order oblivious filtering. This eliminated the intermittent classifications prevalent in the predictions of the original methodology and in some cases improved its accuracy. Stacking aimed to combine the predictions of multiple classifiers to improve prediction rates. Even though it achieved this in some cases, the inclusion of stacking/feature-weighted stacking dramatically improved the reliability of the fault detection model. Finally, a few extra additions were discussed, including the use of a One-Class Support Vector Machine and a Convolutional Neural Network. As a large number of additions have been made, the structure of the offline and online detection methodology implementations has also been defined, providing guideline parameters for each version.

Given faults can be detected, the next chapter discusses a new fault classification methodology. This methodology aims to sit on top of the fault detection methodology and identify the type of a detected fault.

Chapter 6: Fault Classification Method

6.1 Introduction

The previous methodological chapters have focussed on the detection of a fault in sewer surveys. However, detecting a fault is only the first step in sewer survey analysis. Once detected a fault must be labelled, identifying its type according to those outlined in the *Manual of Sewer Condition Classification* (WRc, 2013). This is a more complex task, requiring a detected fault to be categorised into one of 12 distinct fault types. It should be noted that each category also contains many sub-categories. For example, ‘crack’ faults can be broken down into ‘longitudinal’, ‘circumferential’, ‘spiral’ and many other types of ‘crack’ fault. Given the already complex nature of the 12 fault categories, no attempt is made here to identify these sub-types.

The presented fault classification methodology aims to identify the type of a fault within a CCTV frame. This frame would have previously been flagged as containing a fault by a fault detection methodology (see Chapters 4 and 5). Following the aims set out in Section 1.4.2, this tool can sit on top of the defined fault detection methodology (or any other fault detection methodology), improving the detail provided by any automatic analysis. As per the aims, the technology has been designed with flexibility in mind, generalising well across all sewer environments and catering to all fault types within a single streamlined methodology. Similarly, the methods have been developed to perform on or offline, identifying faults during recording or after collection.

As with the fault detection methodology, the method itself relies on computer vision (Section 2.5), applying a collection of Random Forest classifiers to video image feature descriptors diagnosed as containing faults, in order to identify fault types.

This chapter begins by describing the fault classification methodology, outlining its similarities to the fault detection process and highlighting the alternative classification architectures. From here the system is applied to the Wessex Water core dataset, comparing the application of four different classification

architectures on unseen still images. In addition, multi-label classification is also demonstrated as a method for identifying multiple fault types within a single frame. Finally, the most effective ‘1 vs all’ classification architecture is applied to an unseen CCTV sequence, analysing its performance in a real-world environment with the addition of Order Oblivious Filtering (see Section 5.2).

6.2 Fault Classification Methodology

6.2.1 Overview

The fault type identification methodology applies an ensemble of image processing and machine learning techniques to CCTV frames that were identified as containing faults. As in the case of fault detection methodology, this tool is data driven, learning relevant image characteristics from a dataset of pre-labelled images. These labels not only identify the presence of a fault, but also that fault’s type, as per the 12 overarching fault categories.

A breakdown of the fault categories can be seen in Table 3.1. As such, it is assumed that the training database contains a sufficient number of relevant examples of each required fault type. This database could be regularly updated to include images of labelled faults as they are detected in the field. By constantly updating this database and recalibrating the methodology, its effectiveness should continue to improve, much like an engineer gaining surveying experience. Furthermore, the calibration process is entirely automated, requiring no human interaction. As such this can be performed overnight, reducing the impact of the training process (<30 minutes in this case study).

In addition to the many categories of fault, multiple fault types are often present in a single image. These faults are all labelled, although faults are labelled in order of prevalence, starting with the most dominant fault. The base classification methodology only aims to detect this most dominant fault, although as described in Section 6.2.5, multi-labelling can overcome this issue.

When applied to continuous sequences of frames, each frame can be processed in turn, identifying the type of a present fault, as illustrated in Figure 6.1. The methodology’s structure can be intuitively broken down into three stages: ‘Pre-processing’, ‘Feature Extraction’ and ‘Classification’. By design, two of these three stages, ‘Pre-processing’ and ‘Feature Extraction’ are identical to those of

the fault detection methodology, requiring no duplication if both tools are applied in tandem.

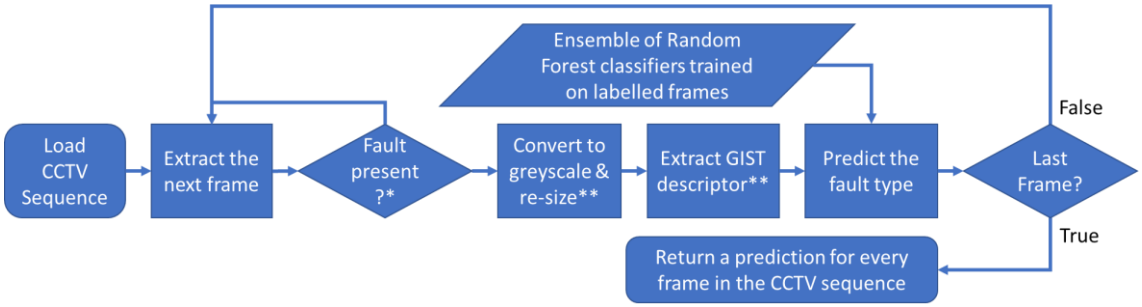


Figure 6.1. *Flow chart defining the structure of the fault identification methodology, when applied to continuous CCTV footage. * As this methodology does not predict the presence of a fault, another methodology should be used for this stage, see Chapter 4. ** If the fault detection methodology defined in Chapter 4 is applied, these steps can be skipped as they are completed by the detection tool.*

6.2.1 Pre-processing & Feature Extraction

Both the 'Pre-processing' and 'Feature Extraction' stages are identical to those presented for the fault detection methodology (see sections 4.2.2 and 4.2.3 for respective details). The choice to use the same features for detection and classification has been made so as to enable an easy integration of the two methods. This integration can be performed by simply re-using the calculated GIST feature vector (Oliva & Torralba, 2001) for each frame from the detection methodology. Alternative feature descriptors, such as HOG (Dalal & Triggs, 2005) and SIFT (Lindeberg, 2012) have been tested in place of the GIST descriptor used here. These alternative descriptors performed, at best, similarly to GIST, whilst taking much longer to process, hence were not ultimately used in the thesis.

6.2.2 Classification

The classification stage aims to identify the fault type in a frame, given its GIST feature descriptor. This identification is performed as per the *Manual of Sewer Condition Classification* (WRc, 2013) into 12 fault classes. Unlike the fault detection methodology, the fault classification method utilises many Random Forest classifiers, arranged in an architecture to perform multi-class classifications. After experimentation it was found that the '1 vs all' architecture

was the most suitable and is discussed here. Details of alternative architectures (which were also tested) are given in Section 6.2.3.

The ‘1 vs all’ architecture classifies a frame’s contents by predicting the probability of each fault type’s presence against all others. This is achieved using a collection of Random Forest (RF) classifiers (Breiman, 2001), one for each fault type. These RF classifiers are calibrated using the ‘Extra-Trees’ algorithm (Geurts, Ernst, & Wehenkel, 2006), which was selected for its good performance in the fault detection method. Given a CCTV frame, each RF attempts to identify its own fault type from the remaining 11 fault classes. This process returns a predicted probability of each fault type’s presence. Once a predicted probability for each fault type is calculated, the faults are ranked from most to least likely. Finally, the highest ranked fault type is assigned to the frame (i.e. identified as the fault type for that frame), completing the classification. This simple approach breaks the multi-class (i.e. 12-class) classification into many (12) simpler binary classifications, improving the effectiveness of the classifiers, which often perform better on these simpler tasks (Essid, Richard, & David, 2006).

Preliminary experiments (not shown here to save space) compared the RF classifier to the Support Vector Machine (SVM) (Cristianini & Shawe-Taylor, An introduction to support vector machines and other kernel-based learning methods, 2000) for use in this classification tool. These showed that the RFs achieved a higher accuracy, and the RF was therefore selected for subsequent studies. This is interesting as the SVM is a binary classifier and is sued in the ‘1 vs all’ configuration for multi-class classification, whereas a single RF is capable of multi-class classification. A full description of the RF and SVM technologies can be found in Section 4.2.4.

6.2.3 Alternative Fault Classification Architectures

The case study presented in Section 6.3 aims to compare three other classification architectures to the ‘1 vs all’ methodology described above. These include:

‘Single’ Random Forest – As Random Forests can perform multi-class classifications (Breiman, 2001), a single random forest was used as a benchmark for the other classification approaches. This requires very little internal adjustment to the RF’s structure,

instead of voting over two classes, decision trees vote over the 12 labelled fault categories. This implementation has the advantage of being the simplest approach, requiring (by far) the least time to process.

'Pairwise' classification – Pairwise classification trains a RF classifier to compare every fault type against every other fault type, resulting in $n^2 - n$ classifiers, where n is the number of fault categories. As each classifier is only making a binary decision between two fault types each class has $n - 1$ associated predictions. These predicted probabilities are then summed to give a score for each fault type, with every fault's score being ranked against the rest. Much like '1 vs all' classification a frame is then assigned the type of the largest score.

'Weighted pairwise' classification – Weighted pairwise classification follows the same architecture as pairwise classification, however each classifier's predicted probability is weighted before being summed to give a score (Hüllermeier & Vanderlooy, 2010). By doing so it is hoped that bias within the dataset can be negated, adding weight to underrepresented fault types and enabling a clearer separation between faults with similar appearances. These weights were learnt using the CMA-ES evolutionary algorithm (Hansen et al. 2003) using a subset of the calibration data unused for training the component classifiers. This algorithm runs during the training process, identifying the weights to be applied to unseen data. The workflow of this technique is presented in Figure 6.2.

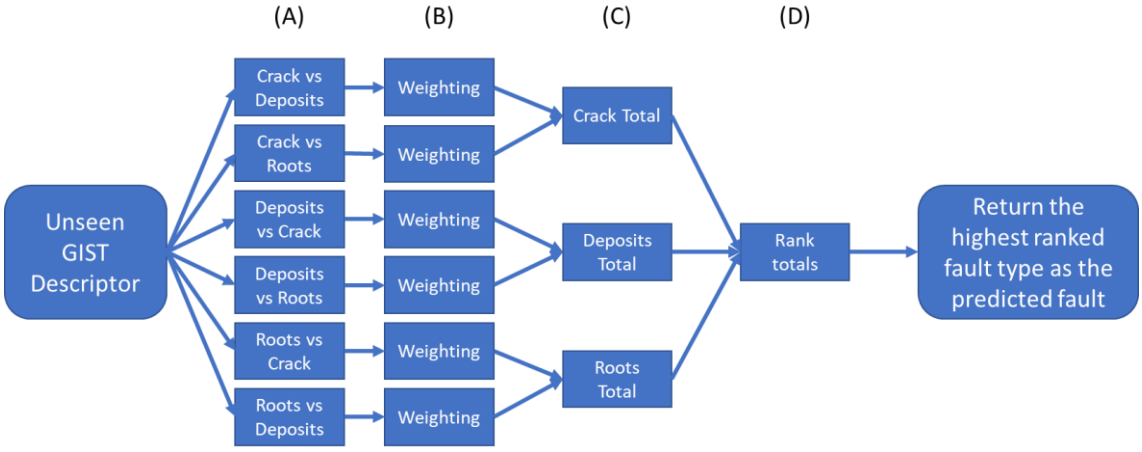


Figure 6.2 *Flow chart demonstrating the application of weighted pairwise classification to determine between 'cracks', 'deposits' and 'roots' fault types. (A) application of each pairwise classifier. (B) multiplying each classifier's predicted probability by its associated weight. (C) Summing the weighted predictions for each fault type. (D) ranking the weighted prediction total for fault types.*

6.2.4 Multi-label Classification

As previously mentioned, it is common for multiple faults to occur together, with a single frame containing many different fault types. Given the currently defined fault classification methodology identifies the most prominent fault, a multi-labelling strategy could identify multiple faults in an image. The chosen multi-labelling strategy utilises the same architecture as the '1 vs all' setup. Implementing a classifier to distinguish each fault type from all others, the classifier's predictions are ranked from most to least probable. However, instead of selecting the highest ranked fault type, a frame is instead labelled to contain any fault with a predicted probability over a given threshold. This threshold should be selected based on the needs of the user, however in this work a default threshold of 0.5 was used.

6.3 Case Study

This case study demonstrates the application of the new fault classification methodology to still images. Using the core Wessex Water dataset (described in Section 3.2), the case study compares the '1 vs all' classification architecture to those outlined in Section 6.2.3. All of these architectures utilise GIST feature descriptors and trained random forest classifiers. Furthermore, the study explores the effectiveness of multi-label classification as outlined in Section 6.2.4. Finally, the study applies the most successful '1 vs all' methodology to a segment of

unseen continuous CCTV footage, evaluating its effectiveness as it would be applied in practice.

6.3.1 Data and Setup

The fault classification methodology was trained using the 2,424 images containing faults from the 7,953 images in the core Wessex Water dataset. These images contained 12 categories of fault, distributed as shown in Table 6.1. In addition, to ease the problems associated with the detection of multiple fault types in a single image, an additional ‘multiple faults’ category was generated. The inclusion of this additional ‘multiple’ category simplified the classification problem, although this is better tackled by multi-label classification discussed later.

As the methodology is applied to still images, all parameters for pre-processing and calculation of GIST descriptors remain the same as those used in the fault detection methodology (Table 5.9). In order to achieve a level of parity throughout the classification architectures, a similar number of trees were used across all RFs. This meant that the single RF classifier used 1,000 trees, ‘1 vs all’ RFs each contained only 100 trees (for a total of 1,300) and both ‘pairwise’ and ‘weighted pairwise’ RFs only contained 10 trees (for a total of 1,560). Even though there is a large discrepancy in trees between classification architectures, by returning to Figure 4.3, it can be seen that the RF performs similarly in binary classifications whether using 10 or 1,000 trees. Like previous case studies a system of 25-fold cross validation was applied to maximise the dataset’s use (Kohavi, 1995) and the generalisation accuracy was calculated across all folds.

Table 6.1 *Distribution of fault types (including the 'multiple' fault type) used in the fault classification case study.*

Fault type	Subtype	Percentage (%)
Joint	Displaced, Open	31.5
Deposits	Attached, Settled	15.7
Multiple	-	12.7
Crack	Longitudinal, Circumferential, Multiple, Spiral	10.0
Surface	-	9.9
Roots	Fine, Tap, Mass	8.2
Infiltration	Running, Gushing	4.8
Obstacles	Intruding Junctions, Masonry, Protrusion	2.8
Other	Vermin, Lining	1.4
Broken / Collapsed	-	1.3
Hole	-	0.9
Brickwork	Missing mortar, Displaced bricks, Missing bricks	0.5
Deformation	-	0.3

6.3.2 Results and Discussion: Still Images

To objectively compare the four classification architectures, this case compares the raw accuracy defined as the percentage of frames correctly classified by each technique. However, this measure alone gives only a naive understanding of each technique's performance. Because of this, the confusion rate matrix for each technique is examined too, to highlight the strengths and weaknesses of each approach. Finally, as each classification technique also ranks fault types from most to least likely in each frame, the accuracy over the most likely 2, 3 and 5 predictions are considered. This last measure could be useful in the development of a detection support tool, working alongside a technician to offer a choice of 2, 3 or 5 most likely faults when the methodology struggles to identify a single fault type. In addition, preliminary tests are performed and analysed using the multi-label classification defined in Section 6.2.4, taking the first steps towards identifying multiple different fault types in a single frame. The most comparable work so far was performed by Hawari et al. (2018) who achieved an accuracy of around 50% when identifying 'crack', 'deposit', 'deformation' and 'joint' faults using an ensemble of techniques.

Once applied to the entire dataset of 2,424 frames and taking the engineer's labels as the ground truth, the 'Single', '1 vs all' and 'Pairwise' techniques performed well (see Table 6.3, 6.4 and 6.5), achieving accuracies of 62.5%, 63.3% and 62.3% respectively. These accuracies indicate that automated techniques can be sufficiently accurate as to aid surveyors in identifying fault types, performing better than randomly guessing a class (7.7%) or always choosing the most common class (31.5%). In addition, by examining Tables 6.3, 6.4 and 6.5, it can be seen that all architectures share a similar pattern of misclassification. On the other hand, the 'Weighted pairwise' classification, achieved a much lower accuracy of 17.7%, struggling to make accurate predictions. When examining the misclassifications in more detail it is clear that this technique's predictions have a strong bias to the 'multiple' fault type misclassifying over 70% of faults as 'multiple' (Table 6.2). Other than predicting the presence of multiple faults, 'weighted pairwise' classification appears to achieve a high classification rate when detecting 'deformation' faults. However, the dataset only contained five examples of this fault, as such there is not sufficient proof to justify this classification architecture. If the number of trees used by each RF classifier within the architecture were to be increased, the architecture's performance may improve, although it is already utilising the most decision trees (1,560 in total). Similarly, the poor performance could be attributed to a lack of data, requiring many more examples to identify suitable weights during optimisation.

Table 6.2 *Confusion rate matrix for 'weighted pairwise' classification. Rates shown are as %.*

'Weighted Pairwise' classification (17.7%)		Predicted Class												
		Surface	Multiple	Deposits	Joint	Infiltration	Obstruction	Roots	Brickwork	Crack	Hole	Other	Deformation	Broken
Correct Class	Surface	**4.9**	70.7	4.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	0.4	0.0	0.0
	Multiple	3.1	**68.4**	5.6	20.5	0.0	2.1	0.0	0.0	0.0	0.0	0.0	0.0	0.3
	Deposits	3.7	70.8	**4.5**	20.2	0.0	0.6	0.0	0.3	0.0	0.0	0.0	0.0	0.0
	Joint	3.9	72.5	3.2	**19.5**	0.0	0.3	0.0	0.0	0.3	0.0	0.0	0.1	0.1
	Infiltration	3.6	72.1	2.7	18.9	**2.7**	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Obstruction	4.9	59.0	0.0	16.4	0.0	**19.7**	0.0	0.0	0.0	0.0	0.0	0.0	0.0
	Roots	3.8	70.4	3.8	19.4	0.0	0.0	**2.7**	0.0	0.0	0.0	0.0	0.0	0.0
	Brickwork	18.2	36.6	9.1	9.1	0.0	0.0	0.0	**0.0**	0.0	0.0	0.0	0.0	0.0
	Crack	6.2	70.8	3.1	3.1	0.0	0.0	0.0	0.0	**2.2**	0.0	0.0	0.0	0.0
	Hole	0.0	65.0	5.0	5.0	0.0	0.0	0.0	0.0	5.0	**0.0**	0.0	0.0	0.0
	Other	3.2	61.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**16.1**	0.0	0.0
	Deformation	0.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**80.0**	0.0
	Broken	0.0	78.6	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**7.1**

Continuing to examine the confusion rate matrices for the more successful techniques, all three architectures performed very similarly (Table 6.3, 6.4, 6.5). All techniques struggled to identify 'multiple', 'obstruction', 'brickwork', 'hole', 'other' and 'broken' fault types achieving True Positive Rates (TPR) less than 50% for each fault type. In the case of the 'multiple' fault type, this outcome is likely due to the large variety of fault combination that the category can include, given this experiment covers 12 distinct categories of fault. The remaining fault categories ('obstruction', 'brickwork', 'hole'. 'other' and 'broken') are likely to have been misclassified due to their poor representation in the dataset, each having fewer than 100 examples, making up less than 5% of the dataset. In all above cases, training on a larger number of examples of faults would likely improve the predictions. By doing so, all of the fault categories would be better representative of the wide variance of each fault's appearance.

Table 6.3 Confusion rate matrix for 'single' RF classification. Rates are shown as %.

'Single' RF classification (62.5%)		Predicted Class												
		Surface	Multiple	Deposits	Joint	Infiltration	Obstruction	Roots	Brickwork	Crack	Hole	Other	Deformation	Broken
Correct Class	Surface	**53.8**	7.1	8.4	26.2	0.0	0.0	1.8	0.0	2.2	0.0	0.4	0.0	0.0
	Multiple	4.2	**31.2**	15.6	32.3	0.3	2.8	5.9	0.0	6.9	0.3	0.0	0.0	0.3
	Deposits	3.1	8.1	**68.3**	18.3	1.1	0.3	0.3	0.0	0.6	0.0	0.0	0.0	0.0
	Joint	2.1	3.5	4.1	**87.2**	0.3	0.0	0.3	0.0	2.5	0.0	0.0	0.0	0.0
	Infiltration	0.0	4.5	2.7	27.9	**63.1**	0.0	0.9	0.0	0.9	0.0	0.0	0.0	0.0
	Obstruction	1.6	23.0	24.6	11.5	4.9	**27.9**	1.6	0.0	4.9	0.0	0.0.0	0.0	0.0
	Roots	2.7	10.2	8.6	20.4	1.6	0.0	**53.2**	0.0	3.2	0.0	0.0	0.0	0.0
	Brickwork	0.0	18.2	72.7	9.1	0.0	0.0	0.0	**0.0**	0.0	0.0	0.0	0.0	0.0
	Crack	2.2	6.6	4.0	27.0	0.4	0.0	0.4	0.0	**58.4**	0.0	0.0	0.0	0.9
	Hole	20.0	20.0	15.0	15.0	0.0	0.0	0.0	0.0	15.0	**15.0**	0.0	0.0	0.0
	Other	6.5	16.1	19.4	25.8	3.2	0.0	0.0	0.0	0.0	0.0	**25.8**	0.0	0.0
	Deformation	0.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**80.0**	0.0
	Broken	0.0	17.9	21.4	39.3	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**17.9**

Table 6.4 *Confusion rate matrix for '1 vs all' classification. Rates are shown as %.*

'1 vs all' classification (63.3%)		Predicted Class												
		Surface	Multiple	Deposits	Joint	Infiltration	Obstruction	Roots	Brickwork	Crack	Hole	Other	Deformation	Broken
Correct Class	Surface	**57.8**	3.6	8.4	24.4	0.9	0.0	1.3	0.0	1.8	0.9	0.9	0.0	0.0
	Multiple	6.6	**30.2**	14.6	31.9	0.7	2.8	4.9	0.3	7.3	0.3	0.0	0.0	0.3
	Deposits	3.7	7.6	**69.4**	16.9	0.3	0.3	0.6	0.3	0.3	0.6	0.3	0.0	0.0
	Joint	2.0	4.5	3.1	**86.1**	0.8	0.0	0.1	0.0	3.1	0.0	0.1	0.0	0.1
	Infiltration	0.0	3.6	0.9	27.0	**65.8**	0.0	0.9	0.0	1.8	0.0	0.0	0.0	0.0
	Obstruction	1.6	21.3	19.7	8.2	4.9	**34.4**	3.3	0.0	4.9	1.6	0.0	0.0	0.0
	Roots	2.7	8.6	8.6	22.0	2.2	0.0	**54.3**	0.0	1.6	0.0	0.0	0.0	0.0
	Brickwork	0.0	18.2	63.6	0.0	0.0	18.2	0.0	**0.0**	0.0	0.0	0.0	0.0	0.0
	Crack	2.2	8.0	3.5	23.0	1.3	0.0	0.9	0.4	**60.2**	0.0	0.0	0.0	0.4
	Hole	10.0	35.0	5.0	15.0	0.0	5.0	5.0	0.0	5.0	**20.0**	0.0	0.0	0.0
	Other	16.1	16.1	9.7	16.1	3.2	0.0	0.0	0.0	9.7	3.2	**25.8**	0.0	0.0
	Deformation	0.0	0.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**80.0**	0.0
	Broken	3.6	28.6	21.4	21.4	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**25.0**

Table 6.5 *Confusion rate matrix for 'pairwise' RF classification. Rates are shown as %.*

'Pairwise' Classification (62.3%)		Predicted Class												
		Surface	Multiple	Deposits	Joint	Infiltration	Obstruction	Roots	Brickwork	Crack	Hole	Other	Deformation	Broken
Correct Class	Surface	**52.0**	7.6	8.4	28.0	0.0	0.0	1.3	0.0	2.2	0.0	0.4	0.0	0.0
	Multiple	3.8	**37.2**	14.9	29.5	0.3	2.8	3.8	0.0	6.9	0.3	0.0	0.0	0.3
	Deposits	2.5	8.4	**66.3**	21.1	0.8	0.3	0.3	0.0	0.3	0.0	0.0	0.0	0.0
	Joint	1.5	4.8	3.4	**87.8**	0.1	0.0	0.1	0.0	2.2	0.0	0.0	0.0	0.0
	Infiltration	0.0	2.7	1.8	32.4	**58.6**	0.0	2.7	0.0	1.8	0.0	0.0	0.0	0.0
	Obstruction	1.6	32.8	18.0	13.1	4.9	**26.2**	3.3	0.0	0.0	0.0	0.0	0.0	0.0
	Roots	3.2	9.7	7.0	24.2	0.5	0.0	**53.8**	0.0	1.6	0.0	0.0	0.0	0.0
	Brickwork	9.1	36.4	54.5	0.0	0.0	0.0	0.0	**0.0**	0.0	0.0	0.0	0.0	0.0
	Crack	1.3	10.6	3.1	29.2	0.0	0.0	0.9	0.0	**54.4**	0.0	0.0	0.0	0.4
	Hole	20.0	35.0	0.0	30.0	0.0	0.0	0.0	0.0	5.0	**10.0**	0.0	0.0	0.0
	Other	6.5	22.6	12.9	32.3	0.0	0.0	0.0	0.0	3.2	0.0	**22.6**	0.0	0.0
	Deformation	0.0	20.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	**80.0**	0.0
	Broken	0.0	42.9	17.9	14.3	0.0	0.0	0.0	0.0	3.6	0.0	0.0	0.0	**21.4**

Given that each classification (i.e. fault type identification) technique ranks faults in order of likeliness, the top ranked faults were explored. As before the 'single', '1 vs all' and 'pairwise' classifiers, performed similarly well achieving a highest accuracy of 86% when considering the most likely three out of thirteen fault categories (Table 6.6). This high accuracy lends the automated methodology to a detection support role, offering a shortlist of fault types for frames the methodology is less sure of.

Table 6.6 Accuracy of all classification architectures when considering the top ranked faults.

Architecture	Accuracy (%)			
	Top 1	Top 2	Top 3	Top 4
Single	62.5	76.5	85.0	93.3
1 vs All	63.3	77.4	85.4	93.8
Pairwise	62.4	77.4	86.0	94.8
Weighted Pairwise	17.7	56.2	76.4	91.3

As experiments have shown that methodology struggles to classify poorly represented fault types, the methodology was applied to a reduced dataset. This dataset includes frames containing only single fault types with at least 100 observations. This led to a reduction of the original dataset to 1,816 frames containing the ‘infiltration’, ‘joint’, ‘deposits’, ‘roots’, ‘crack’ and ‘surface’ fault types only.

As it can be seen from Table 6.7, eliminating underrepresented faults significantly improved the accuracy of the techniques, with ‘1 vs all’ classification achieving the highest accuracy of 74.3%, closely followed by ‘single’ at 73.4% and ‘pairwise’ at 73.0%. As in the case of the full data set, ‘weighted pairwise’ classification performed poorly with an accuracy of just 32.5% and a high bias towards the most common ‘joint’ category. Examining the confusion matrix of ‘1 vs all’ classification for this reduced dataset (see Table 6.7), it is clear that other techniques also suffer with this bias towards the ‘joint’ class, albeit to a lesser extent with most non-‘joint’ faults being commonly misclassified as ‘joints’.

In order to eliminate this bias, the methodology could be trained on a uniform number of each fault type. Preliminary experiments show that this did reduce the bias, although provided minimal improvement in accuracy, as the misclassifications were instead distributed across all classes. By extension, alternative more complex sampling strategies could be implemented. Samples could be chosen proportional to the number of examples, i.e. oversampling minority classes and under sampling majority classes. On the other hand, as the dataset’s fault types are distributed similarly to those found in the wild (Table 6.1), it could be argued that this bias towards the most common fault type is desirable, being representative of faults in the field.

Table 6.7 *Confusion matrix for the '1 vs all' classification technique, when applied to the smaller dataset of well represented fault types.*

'1 vs all' classification (74.3%)		Predicted Class					
		Infiltration	Joint	Deposits	Roots	Crack	Surface
Correct Class	Infiltration	**0.5**	29.7	0.9	0.9	1.8	0.9
	Joint	0.7	**89.7**	4.4	0.4	2.9	1.8
	Deposits	0.8	21.3	**71.3**	0.8	0.8	4.8
	Roots	1.1	24.2	9.1	**60.2**	2.2	3.2
	Crack	1.3	27.9	3.5	0.9	**62.8**	3.5
	Surface	1.3	28.0	7.1	1.8	4.0	**57.8**

Finally, multi-label classification was tested, applying the technique described in Section 6.2.5 to all images of faults. When considering all instance of faults across all 2,424 frames there were 2,560 faults. These additional faults can be attributed to the frames previously labelled as containing multiple faults. Overall the methodology achieved an average accuracy of 67.3% across all faults. Average accuracy is defined as

$$Avg.Accuracy = \frac{1}{n}\left(\sum_{i=1}^{N}\frac{|T_i \cap P_i|}{\max(|T_i|,|P_i|)}\right) \tag{6.1}$$

where T_i is the true set of frame labels and P_i is the predicted set of frame labels for the i^{th} frame. This is superior to the peak accuracy of 63.3% achieved when performing single class classification across all faults. The multi-label performance can be further broken down to identify the true and false positive rates for each fault type as shown in Table 6.8.

Table 6.8 *Breakdown of each fault's classification rate when using multi-labelling.*

Fault type	Number of occurrences	True Positive Rate (%)	False Positive Rate (%)
Joint	850	86.6	23.2
Deposits	448	74.1	7.0
Crack	315	57.1	3.2
Surface	285	51.6	2.8
Roots	270	54.8	1.6
Infiltration	139	59.7	0.7
Obstacles	98	41.8	0.2
Broken / Collapsed	61	26.2	0.4
Other	35	31.4	0.2
Hole	34	17.6	0.2
Brickwork	16	12.5	0.1
Deformation	9	44.4	0.0

The breakdown provided by Table 6.8, clearly demonstrates the discussed issue of bias within predictions, with better represented faults achieving the highest TPR. With an alternative sampling strategy this bias would probably be reduced, leading to a reduction of the number of missed secondary faults. Nonetheless, across all classes the FPR remains low with very few faults being misidentified. Overall, the improved accuracy and additional detail provided by multi-class classification shows promise for application and should supersede single class classification.

6.3.3 Results and Discussion: Continuous Footage

This case applies the best performing classification technique '1 vs all', trained on the full dataset of 2260 frames to a 3-minute sewer survey containing 'roots', 'joint' and 'deposit' faults', each of which spanned multiple frames.

As the methodology was applied to contiguous frames, a little extra information can be gained from neighbouring frames when performing the fault type classification. For example, if the previous frames contain a 'root' fault, it is likely the next frame will also contain a 'root' fault. To take advantage of this spatial continuity, the sequence of predictions is smoothed using Order Oblivious Filtering (Yan, Chakraborty, Misra, Jeung, & Abere, 2012). This technique votes on the fault's type over the faults 50 neighbouring frames (25 frames each side

of the frame in question). This simple smoothing technique has been selected over alternatives due to its previous success when applied to the fault detection problem (see section 5.3).

Over the entire duration of footage, the technique achieved an accuracy of 66.3% on frames labelled as containing a fault which is in line with the accuracies seen in Section 3.3.2. However, in 6 out of 7 faulty segments fault types were correctly identified, a breakdown of which can be seen in Figure 6.3.

Examining Figure 6.3 in detail, faults 2, 4, 5, 6 and 7 were clearly identified, with over 90% of the fault's duration being correctly labelled, where a fault's duration is defined as the number of consecutive video frames the fault appears in. Fault 1 has been less convincingly classified, with only 33% of its duration identified. This is likely due to the low severity of this 'joint' fault, making it hard to distinguish between a normal and displaced joint. Finally, fault 3 has been completely missed, with no frames being correctly identified as a 'roots' fault. It was instead classified as a 'deposit' fault. Factors that could have led to this misclassification, could include its short duration, discrete nature, and the arguable presence of multiple faults. As it is common for faults in sewers to appear in clusters, the presence of multiple faults is a topic of importance and requires further investigation. Suggested strategies for overcoming these issues include the implementation of a multi-labelling strategy. This would label a frame with all faults that lie above a given prediction threshold. To check that the filtering did not negatively impact the detection of fault 1 and 3, the experiment was also performed without filtering, but this achieved inferior results.

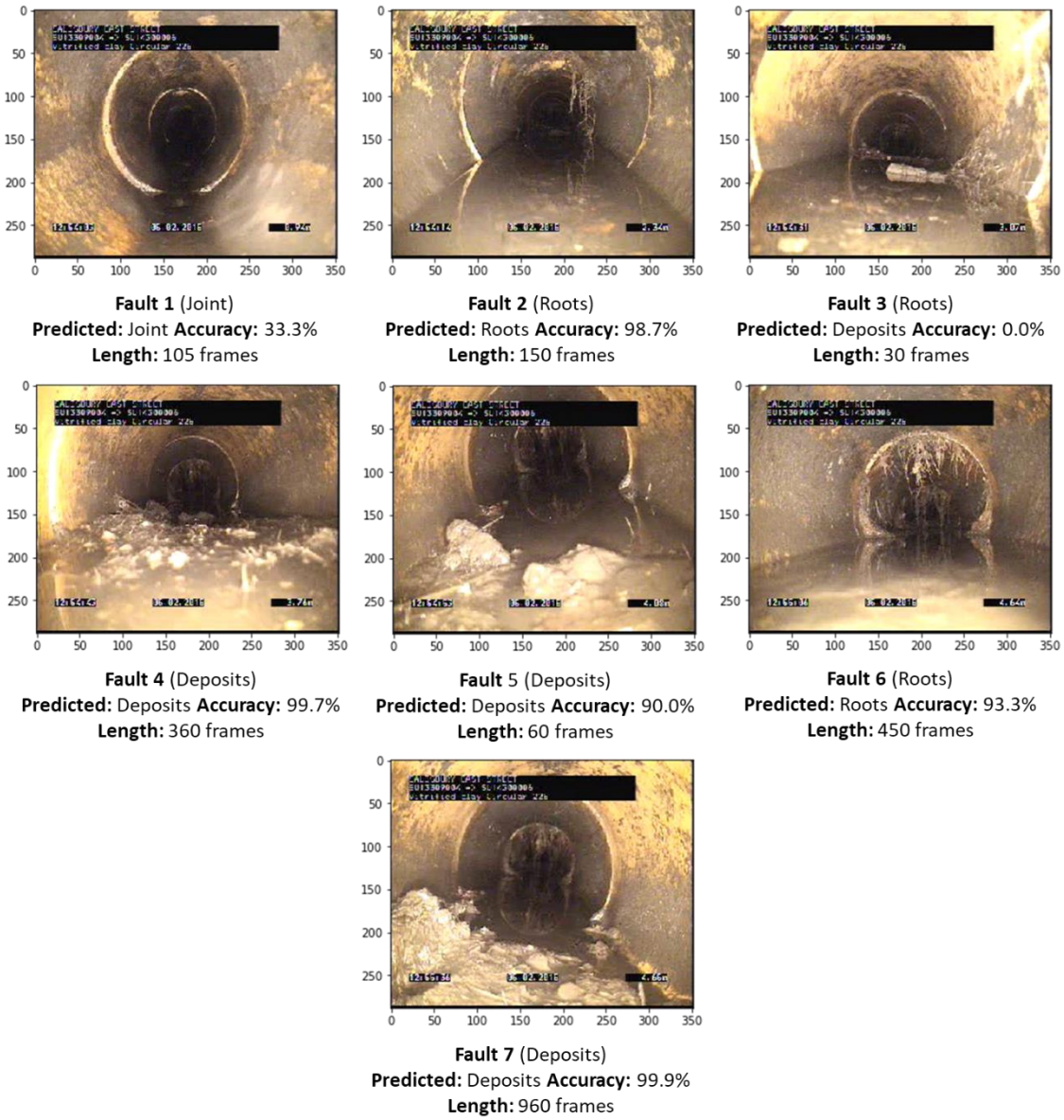


Figure 6.3 *The 7 faults present in the continuous CCTV segment. Labelled with the true fault type, the predicted fault type, the duration of the fault, and the percentage of fault's associated frames correctly identified.*

6.3.4 Summary

In summary, the fault classification methodology can identify the category of a fault within CCTV frames. After comparing multiple classification architectures to perform this task, the '1 vs all' classification method was identified as the best, although 'single' and 'pairwise' achieved very similar results. It achieved an average accuracy of 63.3% across all 12 fault categories and 74.3% when applied to only well represented faults. Furthermore, multi-label classification shows promise for further development and application, achieving an accuracy of 67.3% across all faults.

The importance of the calibration dataset has been evident across all experiments. Well represented faults are often better identified, whilst classifiers tend to incorrectly predict the most underrepresented fault types as the most common class (‘joints’ in the case study shown here). Further research into sampling strategies for the calibration dataset could be the key to eliminating this bias within predictions.

Finally, the ‘1 vs all’ architecture was applied to a continuous CCTV segment, containing a variety of faults. Order Oblivious filtering was also included, to eliminate small inconstancies in predictions. This combination of techniques achieved an accuracy of 66.3% missing only one of the seven faults.

6.4 Summary

In conclusion this chapter presents a new methodology for the classification of faults within an image. A collection of classification strategies has been discussed and demonstrated to identify the ‘1 vs all’ architecture as the best performing and most suitable. This architecture utilises a Random Forest (RF) classifier for each fault type, distinguishing it from all other faults. It achieved an accuracy of 63.3% when applied to still images containing all fault types and a peak accuracy of 74.3% when applied to still images containing faults.

Furthermore, a multi-label classification approach was also investigated. Capable of identifying multiple faults within a frame, the technique achieved an improved accuracy of 67.3% when applied to all 12 fault types. Finally, the ‘1 vs all’ classification was applied to continuous footage with the addition of smoothing, it missed only one out of seven faults in the test segment.

The next chapter discusses the practical implications of combining both the fault detection and classification methodologies. In addition, ideas and preliminary plans for a detection support tool are presented, with the aim to assist engineers in the field.

Chapter 7: A Detection Support Tool

7.1 Introduction

This chapter aims to address the combination and integration of fault detection (Chapters 4 and 5) and fault classification (Chapter 6) methods into current working practices.

Even though guidelines for annotating faults in sewers have been set out in the Manual of Sewer Condition Classification (WRc, 2013) most surveyors across the UK (and the rest of the world) have differing practices. By offering multiple options for implementations of a fault analysis tool, it is hoped that regardless of current practice, a surveyor could incorporate this technology with minimal interruption.

This chapter begins by identifying the workflows of current surveyors, including practical details about the collection and storage of CCTV data. In doing so key stages can be identified to integrate the automated fault analysis methodologies. In addition, fault labelling practices are briefly discussed. Consulting these practices can direct the output of the fault analysis software, suggesting how to best move from a set of predicted probabilities to a set of meaningful fault labels. After identifying the best practices for the integration of the technology, the implementation of the methodologies as part of a software tool are discussed. Providing mock-up examples for both an online and offline detection support tool, useful features are examined and discussed.

7.2 Integrated Fault Analysis

7.2.1 Current practices

From interacting with and talking to multiple water companies in the UK, it is clear that there are the two main approaches to fault analysis: online and offline. However, a third hybrid approach has recently emerged. This is a cloud-based solution, in which surveys are annotated in an office, whilst they are recorded by engineers on site. Doing so enables the survey to be performed uninterrupted, while performing analysis in a comfortable environment by in-house specialists.

Given these three general approaches, there will still be large variations in practice within the industry, so the practicalities of each approach are explored.

As has been mentioned previously within the thesis, online survey analysis is performed during CCTV recording. A surveyor simultaneously operates the camera and annotates faults. Practically this is extremely efficient, with no need to review footage and minimal delay in annotation. However, this is prone to errors and inconsistencies in faults identified. This approach does leave an extremely small window for the additional processing required by the automated analysis methodologies. Furthermore, footage is often annotated using the propriety WinCan software (CD Lab AG, 2018). Given its closed nature this software may provide challenges when integrating the analysis methodologies directly. Finally, unless optimised, it may be a challenge for both the fault detection and classification methodologies to perform in real-time given the hardware limitations of computers on site. Still, if possible and done, this is likely to improve the consistency and reliability of faults identified.

On the other hand, integration with cloud-based systems could be slightly more flexible. Given that CCTV images must be transferred (via the internet or otherwise) to the office in which analysis is performed, there will be a delay on the live feed. Assuming the analyst in the office does not require any live interaction with the surveyor, this delay is not a problem. This means that as latency is unavoidable in these systems, the extra processing time required for near-real time analysis should be more than manageable. Furthermore, as the automated methodology would be running in an office (or on a server) there are likely to be fewer limitations on computational power.

Finally, offline analysis is the most flexible approach, having a diverse selection of implementation options. As recording and analysis are separated into distinct tasks, automated analysis could occur anywhere between the two. Analysis would be performed in an office, implying few restrictions on computational power and much less restriction on running time. As such surveys could be automatically analysed at the analyst's convenience, whether that be in bulk overnight or on an ad hoc basis, 30 mins before analysis. The popular WinCan software would likely still be used for the analysis, posing issues for integration. However, as there are no requirements for hardware integration, alternative bespoke software could be built for semi-automated offline analysis.

7.2.2 Integrating fault detection and classification methods

In order to provide maximum functionality to a detection support tool, both fault detection and classification should be implemented. Doing so is a simple process, as they are designed with easy integration in mind. However, as the tool would only be used as a screening process and all severe faults should be manually examined, reliability of the detection method is key. This isn't to say classification is not important, but as a tool to assist surveyors, the ability to reliably detect severe faults is key to the technology's adoption by industry. The recommended workflow for both online and offline implementation is outlined in Figure 7.1 and 7.2.

Figure 7.1 *Workflow for the integrated offline fault detection and classification methodology.*

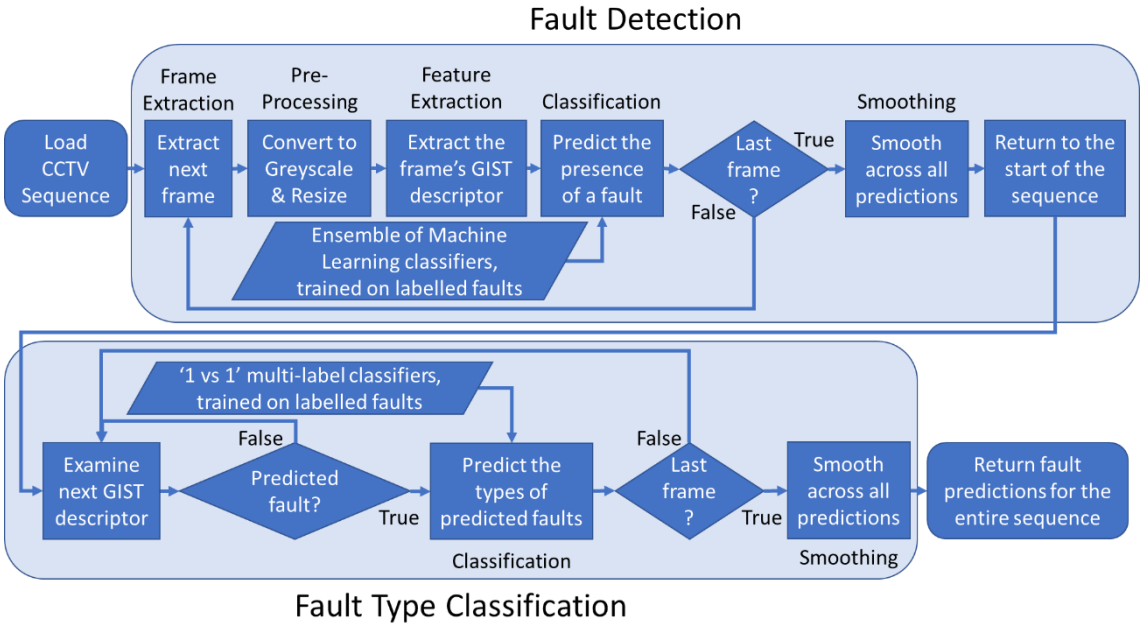
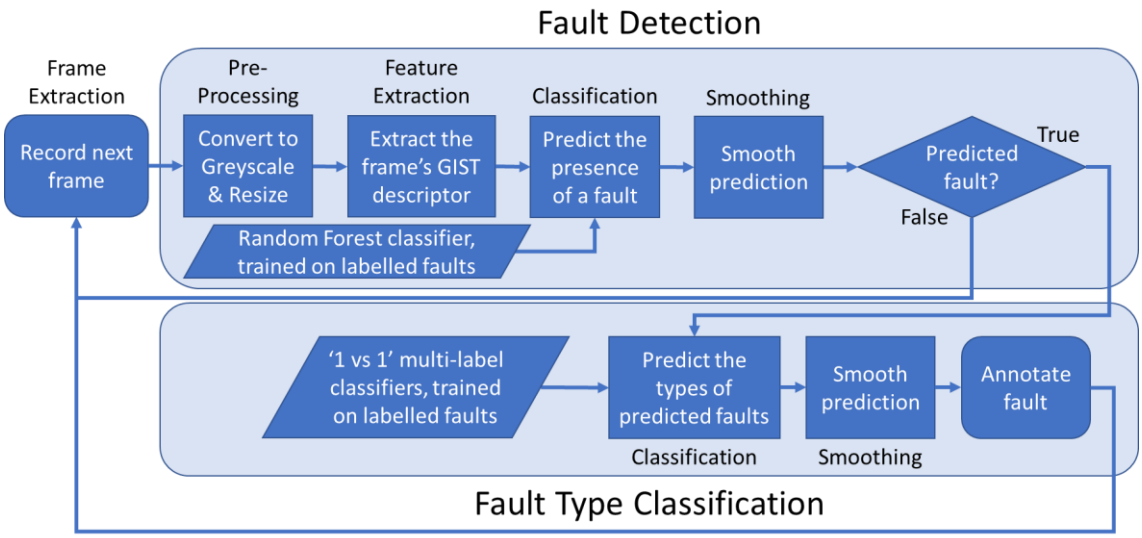


Figure 7.2 *Workflow for the integrated online fault detection and classification methodology.*



Comparing both workflows there are a few differences brought about by the currently restricted processing times of online application, although this may change with future optimised implementations of the methods. Even though the same steps are applied in each, offline implementation passes through the entire sequence of frames twice, applying smoothing after each pass to the entire sequence. Doing so improves the performance increase gained from smoothing, as the sequence can be examined in its entirety. This is much harder to achieve during online implementations as future frames have not been recorded to influence a prediction. Also attributed to time restrictions, 'stacking' is not implemented during online application, as per the notes made in Section 5.7.2.

7.2.3 Labelling practices

The intention when developing both detection and classification methods was to assist engineers in performing effective CCTV analysis. As such a detection support tool would offer suggestions to a surveyor based on its outputs; always leaving the final decision to a human. In their current state, both methodologies are functional but not intuitive, producing a collection of probabilities which must be interpreted by the engineer. To ease implementation the outputs of the techniques should be streamlined for use in a detection support tool. Key considerations include:

The start/end of a fault – Faults are often continuous in nature, with all spanning multiple frames and some spanning the duration of a survey. As such, it can be hard for surveyors to accurately define the start/end of a fault. The

Manual of Sewer Condition Classification (WRc, 2013) states faults should be labelled at the focal length of a camera. Although apparently objective, with variations in camera, the ability to zoom and human interpretation, this is far from true. Hopefully the methods developed in this thesis can bring some uniformity to this problem. Given the calibration datasets should contain annotations from a large variety of surveys labelled by many different surveyors, an average start/end is learned by the machine learning techniques.

Interruptions – Although continuous in nature, faults may not be visible for their entire duration. This could result in the identification of multiple smaller faults, instead of a larger fault which is actually present. PIG devices with pan/tilt functionality would be the biggest culprit of this mislabelling as the camera can be constantly moving. Currently the smoothing stage of the methodology assists in eliminating these isolated inconsistencies, smoothing over a handful of frames where a fault may not be present. However, this will not eliminate larger gaps in annotations, for this reason it is recommended that the use of pan/tilt functionality is limited, and ideally avoided, when recording footage for the detection support tool.

Grouping faults – As the methodologies currently only identify the presence of all fault types within an image, it does not count the number of each fault's occurrence. This means that multiple occurrences of the same fault appearing within an image will only be labelled once. Although not disastrous as the frames will be flagged, the number of instances will still need to be counted by the surveyor. Unfortunately, there is no immediately obvious solution to this problem, which remains a topic for future investigation.

Choosing a decision threshold (thresh.) – As mentioned frequently throughout earlier chapters a key parameter of both detection and classification methods is the choice of decision threshold. Throughout the case studies a threshold of 0.5 has been used, requiring a classifier to predict a frame is more likely to contain a fault than not. Although this usually performs well, the Receiver Operator Characteristic (ROC) curves presented show that this is not always optimal. In order to select a appropriate decision threshold the cost of misclassifications should be estimated which would

allow the decision threshold to be chosen to minimise the overall cost to the water company. It should be noted, however, that these costs may be hard to estimate because in addition to the cost of repair of identified faults (which of course depends on the location of the fault), account should also be made of the cost of failing to repair faults that might develop into more serious and expensive-to-repair faults. In the absence of a thoroughgoing cost analysis, appropriate decision thresholds could be chosen based on engineers' experience and their experience of using the detection support tool.

7.4 Detection Support Tool Implementations

Having discussed some of the practicalities of implementing these technologies, this section provides some suggestions for software implementation. As the needs are clearly different, a design is provided for both an online and offline tool. Each design concentrates on the needs of the surveyor, whilst integrating the predictions of both the detection and classification methodology in an intuitive manner. Neither of the designs have currently been implemented but are presented so as to give a taste of the methodology's future possibilities.

7.4.1 Online Detection Support Tool

The first, simpler interface design is proposed for the online detection support tool. The mock-up shown in Figure 7.3 contains various features and is tailored to the needs of an engineer analysing surveys in the field.

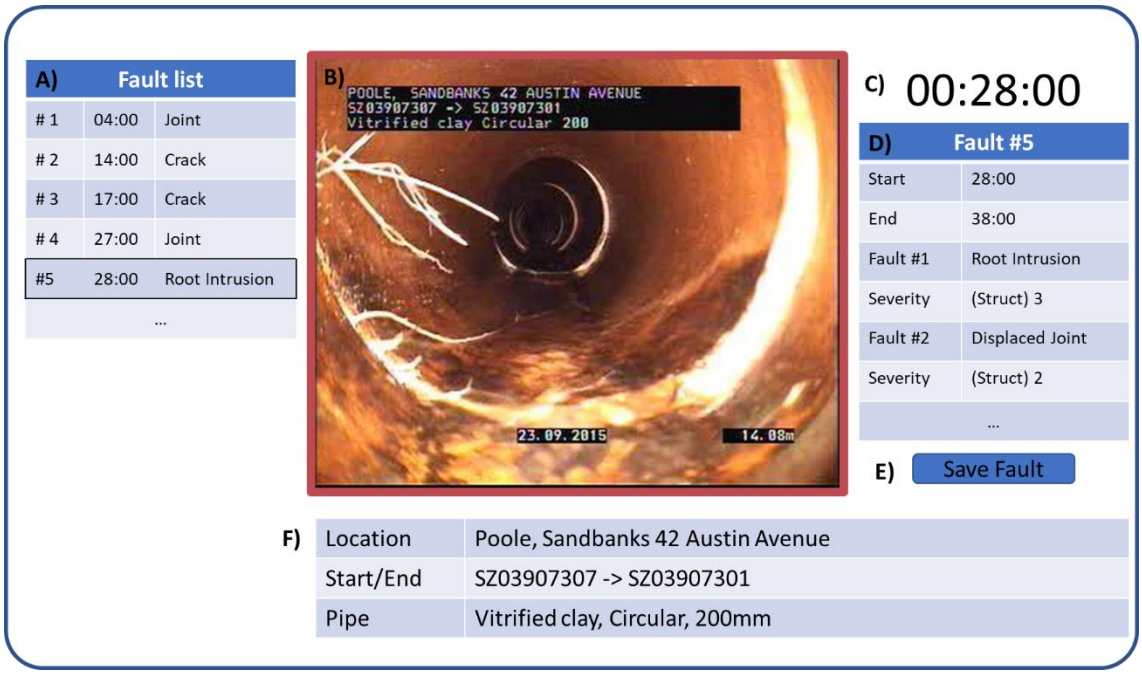


Figure 7.3 *Mock-up of main screen of a detection support tool GUI for automated online analysis*

A description of each of these features is given below:

A) This widget displays all of the annotations previously recorded during the survey. This has been included so as to enable a surveyor to keep track of the current survey and estimate the pipe's state at a glance. It would automatically add annotations to the end of a list as they are saved.

B) The main live display from the surveying camera. The border of the feed can be highlighted (red or green) to signify the presence of a fault. This tells a surveyor that a fault is suspected, without obstructing the image. If fault location were to be implemented (Section 5.6.2), a bounding box could be drawn around faults. However, this may interfere with a surveyor's view of faults and should be a togglable feature.

C) The current video timestamp. This has been included in a large font so as to provide an exact reference point for every annotation. Even though these details are automatically extracted it provides a surveyor with a constant point of reference. An even more useful feature may be the inclusion of a map of the local network, precisely identifying the camera's location.

D) This lists the details of the current fault on screen. The editable list would be where a surveyor inserts the necessary codes and descriptions of a fault according to their surveying practice. However, unlike normal analysis the fault detection and classification methodologies will be able to automatically insert known fault details, for confirmation or editing by the surveyor, thus speeding up the analysis process.

E) Save fault button. Among other functional buttons this would commit the annotation, saving its details ready for the later report.

F) This final box displays the pipe's details. Ideally this would eliminate the need for hardcoding them into an image but have been included for completeness.

The software mock-up (Figure 7.3) is, of course, just one possible interface and it is presented here with an aim to demonstrate how the methods developed in this thesis could be used at some point in the engineering practice. However, many alternatives exist and these would need to be considered in specific applications through a well-defined process of capturing user requirements in the Water industry.

7.4.2 Offline detection support tool

The second, offline interface is slightly more complex, resembling that of video editing software. The mock-up shown in Figure 7.4 contains increased functionality over the online tool, enabling a user to quickly move between faults and scroll through the survey footage.

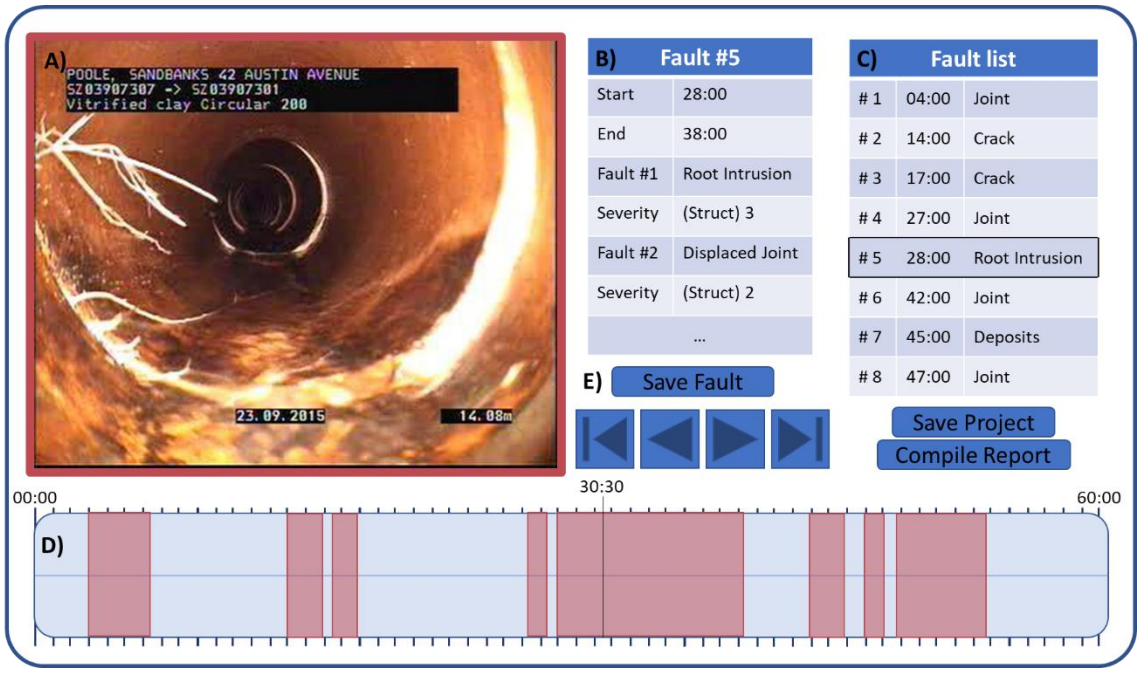


Figure 7.4 *Mock-up of a main screen of a detection support tool GUI for automated offline analysis.*

A breakdown of each of these features is given below:

A) The main display of the offline interface, this shows the current frame of the CCTV survey. As with the online tool, the border is highlighted (red or green) so as to signify the presence of a fault. Again, if fault location is enabled, bounding boxes could be drawn over the image, although this could obscure the view of a fault.

B) The details for all faults within the current frame are given here. This list would be made editable so as to enable a surveyor to insert the necessary details. However, like the online tool the details would be automatically populated by the fault analysis methodologies. This would speed up the analysis process and the predictions could be overwritten by a surveyor if they were to be incorrect.

C) This list details all faults within the current survey, alongside their time stamps. This has been included to provide the surveyor a quick overview of the survey as well as the options for improved navigation. Each of the faults could be linked directly to the video, scrolling to the correct frame upon being clicked.

D) This timeline (showing distance and time) displays the entire duration of the video sequence, much like video editing software. However, on this timeline the presence of a fault is indicated (proportional to its duration) by a red bar. This provides a surveyor with another means of navigating the survey, quickly hopping back and forth along the timeline. Each of these fault bars could be colour coded according to fault type or severity, increasing the information available at a glance.

E) A collection of buttons to enable core functionality within the survey. These can be broken down into navigation and function buttons. Navigation buttons would offer another method of traversing the survey. These could be simple play and re-wind buttons or buttons to skip between points of interest within the survey. Functionality buttons would perform tasks such as saving updates to the current annotation or compiling the standard report necessary for sewer surveys.

Other ideas for features shown by neither mock-up include both the option to view previous surveys at a location, incorporating a database of previous CCTV surveys. Similarly, the derived metric of the overall pipe's condition could be included and dynamically updated. This would enable a surveyor to evaluate the entire pipe's state at a glance. As in the case of the online tool mock up, many alternatives exist and these would need to be considered in specific applications in engineering practice.

7.5 Summary

This chapter discusses the practicalities of implementing both the fault detection and classification methodologies, providing mock-up screenshots of potential software implementations. Current online, offline and cloud-based surveying practices have been discussed, identifying potential opportunities for integration of the fault analysis methodologies. As such combined workflows for online and offline fault analysis have been provided and discussed, highlighting considerations for integration with current working practices.

Finally, examples of mock-up screenshots for both an online and offline fault analysis software have been designed and discussed. These show the potential for the automated analysis technologies, and the scope for integration with current surveying software.

Chapter 8: Conclusion

This thesis has presented a collection of methodologies for automated sewer analysis and demonstrated their application on real CCTV surveys. In this concluding chapter these developments have been consolidated into a final summary. By revising these contributions, key conclusions and thesis contributions are identified. This is followed by recommendations for future developments in this work and the field of automated sewer condition assessment.

8.1 Thesis Summary

The first, introductory chapter of this thesis outlined sewer surveying practices, highlighting the shortcomings of current approaches. In doing so, the scope and aims of the thesis were defined. These aims included the production of automated methodologies capable of detecting and identifying faults using only industry standard CCTV sewer surveys. These methodologies must be able to generalise well across all types of faults and all pipe shapes, sizes and materials. However, most importantly these methodologies must consider the practicalities of real-world application and integrate well with current working practices.

The second chapter explored previous work, relevant to the development of this thesis. This work spanned back to the use of analogue (VHS) recordings, although the application of machine learning was evident from the conception of automated survey analysis. As such, a brief overview of computer vision was also provided, concentrating on the most popular machine learning techniques: Artificial Neural Networks (ANN) and Support Vector Machines (SVM).

The third chapter detailed the data used during the development of the automated analysis methods. Data was provided by multiples sources, however all case studies within this thesis applied the largest and most comprehensive dataset. This was provided by UK water company Wessex Water, and covered over 30km of footage across 613 CCTV videos. This dataset contained a large variety of faults, across a diverse network of sewer pipes. These fault types are broken down in Table 3.1, and according to Wessex Water were typical of their networks.

Chapter 4 was the first methodological chapter, discussing the base fault detection methodology. An overview of the system was provided, breaking it down into its four stages: ‘Frame Extraction’, ‘Pre-processing’, ‘Feature Extraction’ and ‘Classification’. The method made use of GIST feature descriptors and applied either a Support Vector Machine (SVM) or Random Forest (RF) machine learning classifier. To demonstrate the methodology’s effectiveness and compare the use of the SVM and RF classifier, it was applied to still images and continuous video sequences, taken from the Wessex Water dataset. The technique performed well, achieving a peak accuracy of 83% when applied to still images and 85% on video sequences. In both cases these accuracies were achieved by the RF classifier.

Having demonstrated its effectiveness, chapter 5 discussed and demonstrated a collection of improvements to the fault detection methodology. These improvements included ‘smoothing’, which improved accuracy when applied to continuous CCTV videos. It did so by considering neighbouring predictions, eliminating isolated inconsistencies and intermittent predictions which were common in previous demonstrations. A second ‘stacking’ extension was also discussed and demonstrated, in which the predictions of multiple classifiers were combined to boost the techniques accuracy. This was achieved to some extent in the case study, however an equally useful averaging effect was also observed, improving the robustness of the methodology’s predictions. Next a collection of smaller tweaks and alterations were described. These included: identifying the location of a fault within a frame and the application of both the One-Class Support Vector Machine (OCSVM) and Convolutional Neural Network (CNN) classifiers. Finally, these tweaks and amendments were combined to form the structure of an improved ‘extended fault detection methodology’.

Chapter 6 presented the fault classification methodology. This complements the fault detection system, predicting the category of a detected fault. As the two techniques were developed for easy integration, they share the same ‘Frame Extraction’, ‘Pre-processing’ and ‘Feature Extraction’ stages. Unlike detection, the classification method utilises a collection of classifiers to predict the type of a present fault. These were arranged in a ‘1 vs all’ architecture, in which binary predictions are made to distinguish a single fault type from all others. In the case study this architecture was compared to alternatives, achieving the best results

with a peak accuracy of 63% when applied to still images and 74% on continuous CCTV footage, although the ‘single’ and ‘pairwise’ architectures yield very similar accuracies. Given the procedure did not attempt to identify multiple faults of varying types within a single image, multi-label classification was also evaluated using the ‘1 vs all’ architecture. This was capable of identifying multiple different faults within a frame and achieved an average accuracy of 67% when applied to still images.

Finally, chapter 7 discussed the practical application of both fault detection and classification methodologies. The chapter proposes options for integration of both automated survey analysis tools within current online, offline and cloud-based working practices. The chapter continues to discuss key practicalities and potential stumbling blocks for such a system’s implementation, suggesting work-arounds and solutions. Finally, mock screenshots and suggestions were presented for the creation of a detection support tool. These interfaces demonstrate the potential for the technology and its place in future sewer analysis.

8.2 Key Conclusions and Contributions

From this work the key conclusions and main contributions are as follows:

- A novel methodology was developed to determine the presence of sewer faults within CCTV images in an automated way. Unlike in other literature, which develops methods to detect specific fault types, the fault detection methodology presented here is more holistic in nature, i.e. it detects faults of any type using a single method. The new methodology was developed and validated on real life data and has proved to be effective. The methodology has achieved a peak accuracy of 85% when applied to unseen CCTV sequences (see chapter 4 for details).

- Further refinements of above methodology were explored by using additional image processing and machine learning techniques. These techniques include smoothing and stacking, both of which use additional derived information to improve predictions. Smoothing utilises neighbouring predictions to eliminate isolated inconsistencies in video footage, improving the method’s accuracy and reliability. Conversely, stacking combines the predictions of multiple machine learning classifiers

with a secondary stacking classifier in order to utilise the best predictions of each. This has led to further improvements in fault detection accuracy with peak accuracy of 85% achieved when applied to unseen CCTV sequences (see chapter 5 for details).

- A novel methodology for determining the sewer fault type (once detected) was developed. This technique uses an ensemble of Random Forest classifiers to predict the type of a given fault. This differs from other literature, which utilises a collection of dissimilar techniques, engineered solely for detection of individual fault types. The Random Forests are arranged in a '1 vs all' architecture which was shown to be the most effective for the problem. The system also shows promise for application in industry, achieving an accuracy of 74% correctly classified fault types when applied to faults with sufficient training examples (see chapter 6 for details).

- A novel fault analysis system was proposed by combining the fault detection and classification methods developed in this thesis. These methods were combined in an intuitive manner, sharing 'pre-processing' and 'feature extraction' stages so as to streamline the process. In addition, a suitable mock-up of a possible detection support type software tool was provided to illustrate how the combined methodology could be used in engineering practice. It was shown that new methodology could work alongside surveyors, integrating well with current practices to improve the accuracy and efficiency of sewer surveying (see chapter 7 for details).

8.3 Future Work Recommendations

Given the field of automated sewer surveying is still being developed, there is a range of topics that could be considered for additional research. This is further reinforced by the rapid advances in machine learning and the capabilities of new deep learning techniques. More specifically to the methods presented here, there are a few directions of investigation which have been touched upon throughout the thesis and is expanded upon below:

- Both the fault detection and classification methods are still early in the development cycle, as such they would benefit from further testing and

validation on an even larger labelled dataset. Doing so would enable a better demonstration of the fault classification method, which was shown to require at least 100 good quality examples of each fault type. Not only would this improve confidence in the methods, but allow for better tuning of their internal parameters, enabling better guidelines to be set for selecting key decision thresholds. Similarly, the method should be continually compared to alternative machine learning techniques, this is especially prevalent with the continual developments seen in the field of deep learning and the use of Convolutional Neural Networks (CNN).

- Throughout the development of this fault analysis technology, it has been assumed that the codes specified by the *Manual of Sewer Condition Classification* (WRc plc, 2013) are the most appropriate for this project. However, given a standard end user's needs these codes may not be the most suitable for future development. Even though the codes are standard across the UK, less than half of the codes see regular use, with many codes being absent from the 50km of footage provided by Wessex Water. Similarly, many codes are associated with precise measurements which are impossible to accurately measure from video footage alone. The use of a more practical coding system may eliminate the arbitrary precision and subjectivity commonplace in current coding standards. This would compliment the generalised nature of the developed detection support tool, whilst providing the end user with sufficient detail to perform rehabilitation work.

- Although demonstrated in preliminary tests, utilising the Random Forest's (RF) structure to locate faults within CCTV images (section 5.6.2) requires further investigation. Currently only feature importance is utilised to predict a fault's location but given a trees structure additional information could be inferred. Furthermore, as RFs are used in the classification methodology, these could be investigated to identify the location of individual fault types. Finally, this technique should be compared to other methods of object location, the most common of which rely on image segmentation.

- In a similar vein, multi-label fault classification has only been evaluated in small scale experiments. Given it is common for faults to cluster together

in a sewer survey, this technology is key to ensuring the systems usefulness in real world applications. As such this is a topic for continued investigation, aiming to improve the accuracy and reliability of predictions.

- Including additional relevant metadata (i.e. contextual data) when predicting the category of fault, would likely improve the system's accuracy. Although not discussed in detail, a pipe's surrounding environment can greatly increase the chance of certain fault types. For example, a pipe running through a heavily wooded area is likely to experience more root intrusions. Conversely a sewer next to many fast food restaurants may have an increased risk of settled grease deposits. This metadata could include information about the surrounding topology, traffic loadings, pipe material and effluent makeup and could help improve the accuracy of fault type predictions.

- Sampling strategies should be investigated to better select training data for the technique. Most evident in Section 6.3.2, the data used to calibrate the classification methodology relies heavily on a good distribution of labelled examples. The introduction of a sampling strategy is necessary as there is clearly a bias in predictions towards better represented fault types. If this bias were to be eliminated by a sampling strategy the classification method would likely see in improved accuracy.

- Fault severity could be more intelligently calculated, incorporating the knowledge of many surveyors and water companies. Current surveying guidelines prescribe predefined severities to all fault categories. These in turn are used to define the condition of the entire pipe section. However, due to the extreme diversity within a single fault type and the subjective nature of some classifications, these severities are not always accurate. Automated severity analysis could be dynamically achieved using machine learning techniques. As low severity faults are considered to be least interesting to water companies and these are the faults most likely to be missed by the detection method, it would be interesting to correlate severity with detected fault types. Not only would this provide a basis for assigning a condition classification to an entire pipe segment, but it would

also improve a surveyor's trust in the technology, knowing that it will detect the most severe of faults.

- Classification of an identified fault into its subtype would be a logical next step for the technology. Even though overarching fault categories can be determined, a surveyor must still identify the correct subtype and label for any given fault. Alongside this, a standard survey also reports on other factors including the location of assets, the pipe shape/size/material and the condition of manhole chambers. Eventually all of these processes could be automated and incorporated alongside the current system for truly automated sewer analysis.
- As discussed in chapter 7, once the independent automated surveying technologies have been sufficiently developed, they should be combined into a user-friendly detection support tool. This tool could take many forms, whether that be bespoke software or integration with current survey annotation tools. Either way this is a key area of development and research vital to the application of the automated fault analysis technology in industry.

References

Ahray, A., Kawamura, Y., & Ishikawa, M. (2007a*). An automated intelligent fault detection system for inspection of sewer pipes.* IEEJ Transactions on Electronics, Information and Systems, 127(6), 943-950.

Bay, H., Tuytelaars, T., & Van Gool, L. (2006). *Surf: Speeded up robust features.* (pp. 404-417). Berlin, Germany: Springer.

Bovik, A. C., Marianna, C., & Wilson, S. G. (1990). *Multichannel texture analysis using localized spatial filters*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 12(1), 55-73.

Breiman, L. (2001). Random forests. Machine Learning, 45(1), 5-32.

Canny, J. (1986). *A computational approach to edge detection.* Transactions on Pattern Analysis and Machine Inteligence(6), 679-698.

CD Lab AG. (2018). *WinCan - Home*. Retrieved August 6, 2018, from http://www.wincan.com/en/home/

Chae, M., & Abraham, D. (2001). *Neuro-fuzzy approaches for sanitary sewer pipeline condition assessment.* Journal of Computing in Civil Engineering, 15(1), 4-14.

Chaki, A., & Chattopadhyay, T. (2010). *An intelligent fuzzy multifactor based decision support system for crack detection of underground sewer pipelines.* Cairo, Egypt: Intelligent Systems Design and Applications 2010.

Chen, F., Jahanshahi, M., & Joffe, C. (2017). *A texture based methodology using bayesian data fusion for autonomous crack detection on metalic surfaces.* Computer-Aided Civil and Infrastructure Engineering, 32(4), 271-287.

Chen, Y., Zhou, X., & Huang, T. (2001). *One-class SVM for learning in image retrieval.* International Conference on Image Processing, 15(1), 78-83.

Cortes, C., & Vapnik, V. (1995). *Support-vector networks*. Machine Learning, 20(3), 273-297.

Cristianini, N., & Sahwe-Taylor, J. (2000*). An Introduction to Support Vector Machines and Other Kernel-based Learning Methods*. Cambridge University Press.

Dalal, N., & Triggs, B. (2005). *Histograms of oriented gradients for human detection.* Computer Vision and Pattern Recognition, 1, 886-893.

Davies, E. (2004). *Machine Vision: Theory, Algorithms, Practicalities*. Elsevier.

Deb, k., Pratap, A., Agarwal, S., & Meyarivan, T. (2002*). A fast and elitist multiobjective genetic algorithm: NSGA-II.* IEEE Transactions on Evolutionaty Computation, 6(2), 182-197.

Dempster, A., Laird, N., & Rubin, D. (1977). *Maximum likelihood from incomplete data via the EM algorithm.* Journal of the Royal Statistical Society, Series B (Methodological), 1-38.

Duran, O., Althoefer, K., & Seneviratne, L. (2002). *State of the art in sensor technologies for sewer inspection*. IEEE Sensors Journal, 2(2), 73-81.

Duran, O., Althoefer, K., & Seneviratne, L. (2003). *Pipe inspection using a laser-based transducer and automated analysis techniques*. IEEE/ASME Transactions on Mechatronics, 8(3), 401-409.

Duran, O., Althoefer, K., & Seneviratne, L. (2007). *Automated pipe defect detection and categorization using camera/laser-based profiler and artificial neural network.* IEEE Transactions on Automation Science and Engineering, 4(1), 118-126.

Essid, S., Richard, G., & David, B. (2006*). Musical instrument recognition by pairwise classification strategies*. IEEE Transactions on Audio, Speech and Language Processing, 14(4), 1401-1412.

Fawcett, T. (2006). *An introduction to ROC analysis*. Pattern Recognition Letters, 27(8), 861-874.

Gao, W., Zhang, X., Yang, L., & Liu, H. (2010). *An improved Sobel edge detection.* Computer Science and Information Technology, 5, 67-71.

Geurts, P., Ernst, D., & Wehenkel, L. (2006). *Extremely randomised trees.* Machine Lerarning, 63(1), 3-42.

Guo, W., Soibelman, L., & Garrett, J. (2009a*). Automated detection for sewer pipeline inspection and condition assessment*. Automation in Construction, 18(5), 587-596.

Guo, W., Soibelman, L., & J.H., G. (2009b). *Automated defect detection in urban wastewater pipes using invariant features found in video images.* Seattle, USA: Construction Research Congress 2009: Building a Sustainable Future.

Haag, M., & Nagel, H. (1999). *Combination of edge element and optical flow estimates for 3D-model-based vehicle tracking in traffic image sequences.* International Journal of Computer Vision, 35(3), 295-319.

Halfawy, M., & Hengmeechai, J. (2013*). Efficient algorithm for crack detection in sewer images from closed-circuit television inspections*. Journal of Infrastructure Systems, 20(2), 04013014.

Halfawy, M., & Hengmeechai, J. (2014a). *Integrated vision-based system for automated defect detection in sewer closed circuit television inspection videos.* Journal of Computing in Civil Engineering, 29(1), 04014024.

Halfawy, M., & Hengmeechai, J. (2014b). *Automated defect detection in sewer closed circuit television images using histograms of oriented gradients and support vector machine.* Automation in Construction, 38, 1-13.

Hawari, A., Alamin, M., Alkadour, F., Elmasry, M., & Zayed, T. (2018). *Automated defect detetction tool for closed circuit television (CCTV) inspected sewer pipelines.* Automation in Construction, 89, 99-109.

Hearst, M., Dumais, S., Osuna, E., & Platt, J. (1998). *Support vector machines. IEEE Inteligent Systems and Their Applications*, 13(4), 18-28.

Howard, A., Zhu, M., Chen, B., Kalenichenko, D., Wang, W., Weyand, T., .Andreetto, M., & Adam, H. (2017*). Mobilenets: Eficient convolutional neural networks for mobile vision applications*. arXiv preprint arXiv:1704.04861.

Hüllermeier, E., & Vanderlooy, S. (2010). *Combining predictions in pairwise classification: An optimal adaptive voting strategy and its relation to weighted voting.* Pattern Recognition, 43(1), 128-142.

IBAK. (2018). *IBAK - Professional System.* Retrieved August 2, 2018, from https://www.ibak.de/en/produkte/ibak_show/frontenddetail/product/professional-system/

Jahanshahi, M., & Marsi, S. (2012). *Adaptive vision-based crack detection using 3D scene reconstruction for condition assessment of structures.* Automation in Construction, 22, 567-576.

Karpathy, A., Toderici, G., Shetty, S., Leung, T., Sukthankar, R., & Fei-Fei, L. (2014). *Large-scale video classification with convolutional neural networks.* (pp. 1725-1732). Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition.

Khalifa, I., Aboutabl, A., & Aziz, G. (2014*). A new image model for predicting cracks in sewer pipes based on time.* International Journal of Computer Applications, 87(9), 25-32.

Klir, G., & Yuan, B. (1995). *Fuzzy Sets and Fuzzy Logic*. (Vol 4). New Jersey: Prentice Hall.

Kohavi, R. (1995). *A study of cross-validation and bootstrap for accuracy estimation and model selection*. Montreal, Canada: International Joint Conference on Artificial Intelligence.

Krizhevsky, A., Sutskever, I., & Hinton, G. (2012). *Imagenet classification with deep convolutional neural networks.* Advances in Neural Information Processing Systems, 1097-1105.

Längkvist, M., Karlsson, L., & Loutfi, A. (2014). *A review of unsupervised feature learning and deep learning for time series modeling*. Pattern Recognition Letters, 42, 11-24.

LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep Learning*. Nature, 521(7553), 436.

Lindeberg, T. (2012). *Scale invariant feature transform*. Scholarpedia, 7(5), 10491.

Lu, N., Wang, J., Wu, Q., & Yang, L. (2008). *An improved motion detection method for real-time surveillance*. IAENG International Journal of Computer Science, 35(1).

Moselhi, O., & Shebab-Eldeen, T. (1999). *Automated detection of surface defects in water and sewer pipes*. Automation in Construction, 8(5), 581-588.

Moselhi, O., & Shebab-Eldeen, T. (2000). *Classification of defects in sewer pipes using neural networks*. Journal of Infrastructure Systems, 6(3), 97-104.

Murillo, A. (2013). *Localization in urban environments using a panoramic GIST descriptor*. IEEE Trans. Robotics, 29(1), 146-160.

Myrans, J., Everson, R., & Kapelan, Z. (2018b). *Automated detection of faults in sewers using CCTV image sequences*. Automation in Constrauction, 95, 64-71.

Myrans, J., Everson, R., & Kapelan, Z. (2018c*). Automated detection of fault types in CCTV surveys.* Journal of Hydroinformatics, (Accepted for publication).

Myrans, J., Kapelan, Z., & Everson, R. (2016a). *Automated detection of faults in wastewtaer pipes from CCTV footage by using random forests*. Procedia Engineering, 154, 36-41.

Myrans, J., Kapelan, Z., & Everson, R. (2017). *Automatic detection of sewer faults using continuous CCTV footage.* Sheffield, UK: International Computing and Control in the Water Industry Conference (CCWI) 2017.

Myrans, J., Kapelan, Z., & Everson, R. (2018a). *Combining classifiers to detect faults in wastewater networks.* Water Science and Technology, 77(9), 2184-2189.

Myrans, J., Kapelan, Z., & Everson, R. (2018d*). Automatic identification of sewer fault types using CCTV footage*. Palermo, Italy: Electronic proceedings.

Myrans, J., Kapelan, Z., & Everson, R. (2018e). *Using automatic anomaly detection to identify faults in sewers*. International Computing and Control in the Water Industry Conference (CCWI) 2018. 1. Kingston, Canada: WDSA/CCWI Joint Conference Proceedings.

Myrans, J., Kapelan, Z., Everson, R., & Britton, J. (2016b). *Using suport vector machines to identify faults in sewer pipes from CCTV surveys.* Amsterdam, Netherlands: International Computing and Control in the Water Industry Conference (CCWI) 2016.

Oja, T., Pietikainen, M., & Maenpaa, T. (2002). *Multresolution gray-scale and rotation invariant texture classification with local binary patterns*. IEEE Transactions on Pattern Analysis and Machine Intelligence, 24(7), 971-987.

Oliva, A., & Torralba, A. (2001). *Modelling the shape of a scene: A holistic representation of the spatial envelope*. International Journal of Computer Vision, 42(3), 145-175.

Oullette, R., Browne, M., & Hirasawa, K. (2004). *Genetic algorithm optimization of a convolutional neural network for autonomous crack detection.* Evolutionary Computation, 1, 516-521.

Parkhi, O. M., A., V., & Zisserman, A. (2015). *Deep face recognition.* British Machine Vision Conference (BMVC), 1(3), 6.

Platt, J. (2001). *Probabilistic outputs for support vector machines and comparisons to regularized likelihood methods*. Advances in Large Margin Classifiers, 10(3), 61-47.

Pugeault, N., & Bowden, R. (2011). *Driving me around the bend: Learning to drive from visual GIST.* Computer Vision Workshops (ICCV Workshops), 2011 IEEE International Conference on (pp. 1022-1029). Barcelona, Spain: IEEE.

Rabiner, L. (1989*). A tutorial on hidden Markov models and selected applications in speech recognition*. Proceedings of the IEEE, 72(2), 257-286.

Sill, J., Takács, G., Mackey, L., & Lin, D. (2009). *Feature-weighted linear stacking.* arXiv preprint arXiv:0911.0460.

Simonyan, K., & Zisserman, A. (2014). *Very deep convolutional networks for large-scale image recognition.* arXiv preprint arXiv:1409.1556.

Sinha, S., & Fieguth, P. (2006*). Neuro-fuzzy network for the classification of buried pipe defects.* Automation in Construction, 15(1), 73-83.

Steinwart, I., & Christmann, A. (2008). *Support Vector Machines*. Springer Sience & Business Media.

van der Steen, A., Dirksen, J., & Clemens, F. (2014). *Visual sewer inspection:detail of coding system versus data quality?* Structure and Infrastructure Engineering, 10(11), 1385-1393.

Vedaldi, A., & Fulkerson, B. (2010). *VLFeat: An open and portable library of computer vision algorithms*. (pp. 1469-1472). Firenze, Italy: ACM.

WRc. (2013). *Manual of Sewer Condition Classification* (5th Revised Edition ed.). WRc Publications.

Yacoob, Y., & Davis, L. (1996). *Recognizing human facial expression from long image sequences using optical flow.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 18(6), 636-642.

Yan, Z., Chakraborty, D., Misra, A., Jeung, H., & Abere, K. (2012). *Semantic activity classification using locomotive signatures from mobile phones.* Distributed Infromation Systems Laboratory (No. EPFL-REPORT-174016).

Yegnanarayana, B. (2009). *Artificial Neural Networks*. PHI Learning Pvt. Ltd.